Dynamics of Biomedical Networks:

Applications to Gene Regulatory and Large Scale Functional Brain Networks

By

SIMON GODWIN

Department of Engineering Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

DECEMBER 2019

Word count: 31,259

Abstract

Biological systems are complicated, formed of many interacting parts. Network science is well placed to attempt to elucidate these systems, and it is important to further our understanding to look at different biological networks across scales and applications. For this thesis, we examined two biological systems that can be described with networks: gene regulatory networks (GRN) and functional brain networks. We developed and analysed a novel GRN describing pluripotent mouse embryonic stem cells (mESCs) dynamics. Pluripotency is controlled by a complex GRN; understanding the interplay between the network elements and factors present in the different media cultures could help define optimal culture protocols. We expanded an existing GRN describing Nanog (a master regulator of pluripotency) dynamics to include additional genes and both chemicals present in ground-state pluripotency media, i.e. 2i+LIF. We showed, using stochastic differential equations and bifurcation analysis, that the experimental dynamics of Nanog result from the combination of feedback loops in the GRN and transcriptional noise, giving rise to bistability. The functional brain network analysis we performed was in exploring Multiple Sclerosis (MS), a demyelinating disease of the central nervous system. This damage does not correlate with impairment, suggesting the brain can undergo neuroplasticity to compensate. We used functional magnetic resonance imaging (a technique to indirectly measure activity in the brain) collected in Bristol (CRICBristol) to explore neuroplasticity. We preprocessed the scans, extracted time series from regions of interest (ROI) and measured the functional connectivity (FC) between each ROI time series. We abstracted the brain to a network, where nodes are different regions of the brain and edges represent the FC between them. We measured different properties of the networks and compared them between MS and healthy controls (HC). We found no evidence of neuroplasticity between MS and HC, which could be because the MS cohort brains have already successfully undergone plasticity to compensate for the damage caused. Future work for the GRN network could be in using it to understand how chemicals known to interact with its elements could be applied to control pluripotent behaviour of mESCs. For the functional brain networks, a dynamic causal modelling (DCM) analysis could be done to elucidate if there are any differences in FC between MS and HC.

Dedication and acknowledgements

I would like to first and foremost thank both my supervisors for their patient support and teaching me so much. They have both been invaluable in guiding me and in their special ability of finding the mistakes in my work. A quick thanks goes to the EPSRC for funding me during these years, including the exotic conferences abroad. Thank you to my systems biology collaborators, without which we would not have been able to publish our paper. Further thanks to my collaborators in CRIC for their help and guidance on the brain, but also for letting me play with their expensive scanner. Thank you to my fellow PhD students of the Buncaer, especially for some of your input into my Kalman filter chapter. I must give a thanks to my landlord for giving me cheap rent the latter part of my PhD. Thanks to my girlfriend for her love and motivation, and to her family for entertaining me this entire time. To anyone not specifically mentioned, I apologise and give my sincerest thanks. I would finally like to thank my family, for being behind me every step of the way. Extra special thanks go to my mum for her unconditional love and support.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... [redacted] ..... DATE: 15 / 12 / 2019

TABLE OF CONTENTS

Page

List of Tables **xi**

List of Figures **xiii**

1 Introduction **1**

2 Pluripotency Gene Regulatory Networks in mouse Embryonic Stem Cells **5**

2.1 Biology background 5

2.1.1 Culture conditions 6

2.1.2 Pluripotency genes in different culture conditions 6

2.2 Modelling 8

2.2.1 Gene regulatory network (GRN) model 8

2.2.2 Other mechanisms to model Nanog dynamics 9

2.2.3 Reference model 9

2.2.4 How to model a GRN 11

2.2.5 Bistability in an idealised 'core' GRN 18

2.2.6 Parameter effects on bistability in the idealised 'core' GRN 21

2.2.7 Stochasticity to encompass complex molecular processes 24

2.2.8 Interplay between noise strength 'σ_N' and maximum transcription rate of Nanog 'S_N' 25

2.3 Extending the reference model 26

2.3.1 Types of experimental data used 27

2.3.2 Additional genes 28

2.3.3 Gene interaction directions 28

2.3.4 Medium signalling components and protein-protein interactions 29

2.3.5 Putting it all together 29

2.3.6 Summary of final network 31

2.3.7 Parameter values and fitting 32

2.3.8 Simulation and bifurcation analysis 35

2.4 New model behaviour/results 36

2.4.1 Serum+LIF results . . . 36
2.4.2 2i+LIF results . . . 41
2.4.3 Gene/molecular pathway deletions . . . 47
2.4.4 Alternative network structures/modifications to the GRN . . . 50
2.4.5 Alternative simpler networks capable of reproducing bistability in Nanog 58
2.4.6 Agent-based model . . . 65
2.5 Discussion . . . 70

3 Background of fMRI Analysis: A Systematic Review of Resting-State Studies in Multiple Sclerosis 73
3.1 Introduction . . . 73
3.2 Materials and methods . . . 75
3.3 Results . . . 76
3.3.1 Quality assessment . . . 76
3.3.2 MS phenotypes and diagnostic criteria . . . 77
3.3.3 Study aims and size . . . 78
3.3.4 Data acquisition and analysis methods . . . 78
3.3.5 Neuroplasticity in MS . . . 79
3.3.6 Early MS . . . 83
3.3.7 Comparing MS to HC . . . 83
3.3.8 Correlations with clinical markers . . . 83
3.4 Discussion . . . 84
3.5 Conclusion . . . 87

4 Analysing Large Scale Functional Networks in Multiple Sclerosis 103
4.1 Background of study . . . 104
4.2 Study participants . . . 105
4.3 Preprocessing . . . 105
4.3.1 SPM preprocessing . . . 105
4.3.2 Further preprocessing in MATLAB . . . 110
4.3.3 Equivalence of linear regression and time series extraction ordering . . . 111
4.4 Regions of interest (time series extraction) . . . 115
4.5 Network analysis methods . . . 116
4.5.1 Network measures . . . 118
4.5.2 Individual basal ganglia region analysis and single threshold permutation correction . . . 121
4.6 Network analysis results . . . 122
4.6.1 Individual basal ganglia region results . . . 123
4.7 Correlations with clinical markers . . . 125

4.8 Preprocessing protocol alterations 126
4.8.1 Nuisance regression alteration 126
4.8.2 Bandpass parameter space 127
4.9 Discussion 132

5 A Method of Characterizing Dynamic Functional Connectivity in fMRI 135
5.1 Methods 136
5.2 Results 139
5.3 Discussion 142

6 Final Conclusions 145

Bibliography 149

List of Tables

Table		Page
2.1	Experiment types for extending GRN	27
2.2	CODEX	29
2.3	NIA bank	30
2.4	Steady-state concentrations	35
2.5	Deterministic fold change	35
2.6	Full network parameter values	36
2.7	Simple network 1 parameter values	59
2.8	Simple network 2 parameter values	62
2.9	Simple network 3 parameter values	64
3.1	Search terms used for systematic review	75
3.2	ICA table	88
3.3	Seed to voxel table	92
3.4	ROI table	94
3.5	AAL table	95
3.6	ALFF table	97
3.7	Voxel to voxel table	98
3.8	Study demographics	99
3.9	Quality assessment table	102
4.1	MS study participant demographics	105
4.2	T-test results - network measures	123
4.3	T-test results - network measures; including first time derivative of nuisance signals	127

List of Figures

Figure **Page**

2.1 Development of blastocyst. . . . 6
2.2 Medium images . . . 8
2.3 Herberg results . . . 10
2.4 Herberg results . . . 11
2.5 TF binding activation . . . 12
2.6 TF binding inhibition . . . 14
2.7 TF binding activation homodimer . . . 15
2.8 Simple GRNs . . . 18
2.9 Simple core networks . . . 19
2.10 Deriving bistability of Nanog . . . 21
2.11 Sos parameter change . . . 22
2.12 Sn parameter change . . . 23
2.13 K parameter change . . . 23
2.14 Hill coefficent parameter change . . . 24
2.15 Sn and k parameter change . . . 24
2.16 Change of NH proportion with noise and S4 . . . 25
2.17 Process diagram . . . 26
2.18 Final network schematic . . . 31
2.19 Serum parameter results . . . 37
2.20 Serum sorting results . . . 38
2.21 Individual cell time lapse simulations in Serum+LIF (PD=CH=0) . . . 39
2.22 Individual cell time lapse data, Serum+LIF . . . 40
2.23 2i parameter results (PD=CH=20) . . . 41
2.24 2i parameter results (PD=CH=2) . . . 42
2.25 2i sorting results . . . 43
2.26 Individual cell time lapse simulations in 2i+LIF (PD=CH=20) . . . 44
2.27 Individual cell time lapse simulations in 2i+LIF (PD=CH=2) . . . 45
2.28 Indivudal cell time lapse data, 2i+LIF . . . 46
2.29 2-par gene deletions . . . 47

2.30 Prdm14 deletion distributions . . . 48
2.31 β-catenin deletion continuations . . . 49
2.32 Tcf3 deletion continuations . . . 49
2.33 β-catenin activation by Nanog added . . . 50
2.34 Nanog and β-catenin distributions with no activation by Nanog, Serum+LIF . . . 51
2.35 Nanog and β-catenin distributions with activation by Nanog, Serum+LIF . . . 52
2.36 Nanog and β-catenin distributions with no activation by Nanog, 2i+LIF . . . 52
2.37 Nanog and β-catenin distributions with activation by Nanog, 2i+LIF . . . 53
2.38 β-catenin interaction on Mycn removed . . . 54
2.39 β-catenin interaction on Mycn removed, Serum+LIF results . . . 55
2.40 β-catenin interaction on Mycn removed, 2i+LIF results . . . 55
2.41 β-catenin interaction on Mycn activation . . . 56
2.42 β-catenin interaction on Mycn activation, Serum+LIF results . . . 57
2.43 β-catenin interaction on Mycn activation, 2i+LIF results . . . 57
2.44 Simplified network 1 GRN . . . 58
2.45 Continuation of steady-state of Simplified network 1 . . . 60
2.46 Simplified network 2 GRN . . . 61
2.47 Continuation of steady-state of Simplified network 2 . . . 62
2.48 Simplified network 3 GRN . . . 63
2.49 Continuation of steady-state of Simplified network 3 . . . 65
2.50 Nanog cell-cycle fold change fit . . . 67
2.51 Nanog distributions Nanog cell-cycle . . . 67
2.52 Mycn cell-cycle fold change fit . . . 68
2.53 Nanog distributions Mycn cell-cycle . . . 69
2.54 Agent Sorting Nanog . . . 69
2.55 Agent screenshots . . . 70

3.1 Schematic of search term groups . . . 75
3.2 Diagram of review steps . . . 77

4.1 SPM preprocessing realignment . . . 106
4.2 SPM preprocessing segmentation . . . 107
4.3 SPM preprocessing normalisation . . . 108
4.4 SPM smoothing . . . 109
4.5 Extracted time series . . . 111
4.6 Column space . . . 113
4.7 AAL atlas . . . 116
4.8 BOLD vs BOLD plots . . . 117
4.9 Basal Ganglia region by region t-test . . . 124

4.10 Random permutation distribution of basal ganglia regions 125
4.11 Bandpass parameter space - Clustering Coefficient 129
4.12 Bandpass parameter space - Weighted Small-World Propensity 131
4.13 Bandpass parameter space - AAL - Basal inner 132

5.1 Kalman Filter Representation 136
5.2 Dynamic correlation window 137
5.3 Bayesian update of distributions 138
5.4 Distributions of BOLD values 139
5.5 Dynamic Correlation - Window Size 10 140
5.6 Dynamic Correlation - Varying Window Size 141
5.7 End of Kalman process 142

CHAPTER 1

INTRODUCTION

Biological systems are complex, often comprised of many intricate and interacting parts. For example, there are approximately 20,000 genes in the human body [155]. These all interact with each other as well as molecular pathways to create the body and maintain the homeostatis of various organs. Another example is in the human brain which contains approximately 100 billion neurons [63] with many more connections between them, giving rise to the control centre of the body vital for its survival.

The complexity of these systems can lend naturally to modelling them as networks, where the nodes and edges can have distinct biological meanings. In biological systems it is important to consider them as a whole, as well as account for temporal dynamics of the variables of interest. With this understanding, new treatments for diseases, or protocols for *in-vitro* research, might be better informed. Mathematical modelling is a useful tool, which can be used to represent biological networks, complementing physical experiments and allowing a quantitative understanding of the system dynamics. We studied two different biological networks. The first was a network describing interacting genes in stem cells; we explored models of different *in-vitro* growing conditions and analysed how this could change the expression of the genes being modelled and, ultimately, stem cell phenotype. The second network involved the communications within the brain; we investigated how network properties of these functional networks might differ between different populations of interest.

In Chapter 2 we looked at how interactions among proteins, forming Gene Regulatory Networks (GRNs), can determine the fate of mouse embryonic stem cells (mESC) and their propensity to differentiate into specialized cell types. mESCs have a property called pluripotency, meaning

they can divide and give rise to other cells still able to differentiate into any cell type. These cells, isolated from mice, can be indefinitely cultured in the lab to study development *in-vitro*, where significant research effort has been put into defining optimal culture conditions for pluripotency maintenance. The most notable pluripotency gene is Nanog [20], regarded as a master regulator of pluripotency. Nevertheless, mESCs can respond differently to culture media and express different concentrations of genes important for pluripotency. Cells expressing high concentrations of Nanog, Nanog high (NH), posses greater pluripotency and therefore less propensity to differentiate. The opposite is true for Nanog low (NL) cells. Additionally, mESCs are capable of switching between NH and NL states due to temporal fluctuations of Nanog [4], with similar behaviour observed in other pluripotency factors. These cells can be cultured in media [145]: Serum+LIF and 2i+LIF, where Serum+LIF containins serum factors and leukaemia inhibitory factor (LIF) and 2i+LIF is more recently established, replacing serum factor with two inhibitors called PD and Chiron. In these media, mESCs of a previously proposed cell line type termed TNGA display heterogeneous levels of Nanog in Serum+LIF, which when cultured in 2i+LIF is lost [159], replaced by an observed pure NH distribution. This 'ground-state' of pluripotency has recently been challenged, with a newer cell line type, termed ND [3], not displaying an elimination of the NL phenotype in 2i+LIF.

We focussed on the temporal dynamics of Nanog, combining modelling and experiments to understand how its heterogeneous expression and temporal dynamics depend on the culture media and regulate both cell pluripotency and cell division. Many modelling types and mechanisms have been put forward to explain heterogeneity of pluripotent genes. We extended an existing GRN for pluripotency, and used stochastic differential equations (SDEs) to describe interactions of core pluripotency genes and their dynamics. The mathematical model successfully recapitulated the heterogeneous levels of Nanog observed experimentally, and explained them through the combination of feedbacks in the GRN, transcriptional noise and the chemicals present in the culture media. The work in this chapter has been published in Nature Partner Journals Systems Biology and Applications [55].

Another biological system we looked at was based on functional brain networks in Multiple Sclerosis (MS), a demyelinating disease of the central nervous system. MS is a leading cause of disability in young adults and a major health burden [17, 80]. We were interested in what compensatory mechanisms the MS participants brains may undergo to limit the effects of the demyelination, which could inform on disease progression or the effectiveness of treatments.

Chapter 3 is a systematic review, a type of literature review used in clinical settings. The review focused on functional magnetic resonance imaging (fMRI) which measures oxygenated blood flow around the brain, called the blood oxygen level dependent (BOLD) signal. This signal can be

used as a proxy for brain activity, as more active regions of the brain require more oxygenated blood. In particular, the review looked at studies analysing MS with resting-state fMRI, a type of fMRI paradigm that measures basal connectivity and has the advantage of requiring no task or paradigm that could be arbitrary or subject to bias. The review looked at what evidence in the literature shows neuroplasticity in the brains of MS to compensate for their damage.

For Chapter 4 we carried out a study of data obtained from MS participants and healthy controls (HC). We performed resting-state fMRI scans on all the participants, and applied network analysis techniques to their BOLD time series. The aim was to determine if there were any differences between MS and HC which could be evidence of neuroplasticity, i.e. the brain forming new connections to compensate for the damage. Specifically we were looking for evidence which could explain the increased fatigue experienced by the MS cohort. This chapter outlines the process of preprocessing applied to the fMRI data, and the measure of functional connectivity (FC) between distinct predefined regions of the brain. FC is a statistical measure of correlation between the BOLD signal of two regions, with higher FC inferring greater communication between them. The brain can then be abstracted to a network, with regions of the brain being the nodes and the edges between them being the respective FC. We make standard network measurements of these functional brain networks like clustering coefficient, small world propensity and modularity. We then compared the measurements between the two populations to see if any differences occurred, signifying possible compensatory neuroplasticity.

Chapter 5 is a preliminary investigation of a tool to analyse dynamic FC in fMRI. In this analysis we developed the application of a Kalman filter to dynamic FC in an attempt to track and characterize the broad changes to FC over time. The filter behaved like a smoother, broadly tracking FC over time.

CHAPTER 2

PLURIPOTENCY GENE REGULATORY NETWORKS IN MOUSE EMBRYONIC STEM CELLS

Stem cells are a special type of cell which have the ability to turn into other types of cell, after which they may perform specialized functions needed for an organism to survive. A mouse embryonic stem cell (mESC) is a type of stem cell found in the developing embryos of mice which can differentiate into all the cells needed for the embryo to develop into a body. Significant research has been performed on mESCs to define optimal ways in which they can be maintained *in-vitro* into the pluripotent state, and to identify protocols in which their differentiation can be controlled. This could pave the way for future research which could use these cells for regenerative medicines, replacing lost or damaged cells the body cannot repair by itself.

The work in this chapter was published in Nature Partner Journals Systems Biology and Applications [55], where the results of the work I performed are in Sections 2.4.1, 2.4.2, 2.4.3, 2.4.4 and 2.4.5. The work behind the results of Section 2.4.6 were performed by collaborators.

2.1 Biology background

When a mouse egg cell is fertilized a single cell Zygote is formed. This cell continuously divides forming a mass collection of cells after 3-5 days called a blastocyst. Figure 2.1 shows an illustration of a single fertilized egg cell forming into a blastocyst. The blastocyst is broadly formed of two cell types; those which form its outer wall structure, and a group of cells inside called the inner cell mass. The inner cell mass is where mESCs are taken from.

mESCs can differentiate into almost any type of cell in the body. However once a mESC has turned into, for example, a skin cell, it remains this new cell type. A mouse embryo needs

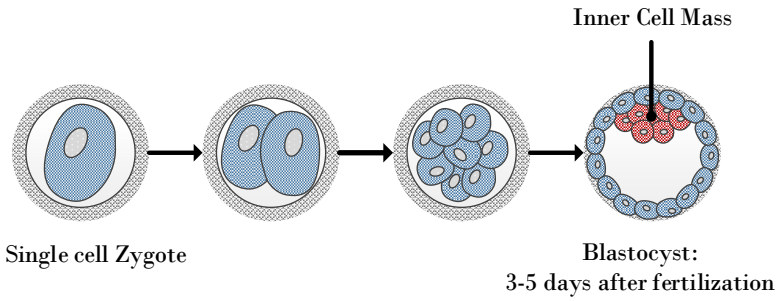


Figure 2.1: Illustration showing the formation of a blastocyst where mESCs are taken from the Inner Cell Mass.

a plentiful supply of mESCs, and the ability to create more of them in order to form all the necessary components of its body. To achieve this mESCs have a property called pluripotency; this is where mESCs have the ability to turn into most other cell types. Additionally with self-renewal, mESCs can split, not into new cell types but more mESCs, and all the mESCs under the right conditions still retain the ability to differentiate.

In-vivo, mESCs in an embryo are pluripotent as the organism has evolved to be able retain the right mESC behaviour and switch it on and off when necessary. Studying mESCs *in-vitro* requires the correct conditions for successful growing and studying of these cell populations.

2.1.1 Culture conditions

In-vitro, mESCs are grown in culture media; culture conditions can confer different characteristics to cell populations. Two media of interest are Serum+LIF and 2i+LIF. Serum+LIF is the more established culture containing serum factors and the cytokine leukaemia inhibitory factor (LIF). The more recently developed medium 2i+LIF [83] also contains LIF but with two chemical inhibitors (2i stands for 2 inhibitor) instead of serum factors. The inhibitors are MEK inhibitor PD0325901, hereafter named PD, and glycogen synthase kinase-3 (Gsk3) inhibitor CHIR99021, hereafter named Chiron (also sometimes referred to as CH).

Cells grown in Serum+LIF have a flat morphology whereas in 2i+LIF the colonies grow in more spherical, contained structures [145]. As well as the morphology of the colonies, mESCs have a number of different characteristics between the two media. These include the differential expression of genes involved in pluripotency, cell-cycle and cell death.

2.1.2 Pluripotency genes in different culture conditions

Genes are units of inheritance, passed down through the generations. They code for proteins which then can perform molecular tasks within cells. When mESCs are cultured in Serum+LIF

the population has a heterogeneous expression of genes which govern pluripotency. The most notable of which is Nanog, a master regulator of pluripotency and development [20]. Cells expressing high concentration levels of Nanog, the Nanog high (NH) phenotype, display greater pluripotency and less propensity to differentiate [138]. Nanog low (NL) mESCs have reduced pluripotency, and therefore a greater propensity to differentiate [138]. Nanog and other pluripotency factors display temporal fluctuations, with single cells switching in time between high and low states. These factors are also heterogeneous, with many of them, directly or indirectly, regulated by Nanog, including Rex1, Stella, Essrb and β-catenin [147].

mESCs need to be pluripotent when cultured in pluripotency conditions, while keeping the ability to differentiate when differentiation stimuli are provided. For this aim, a more uniform population would have all mESCs expressing the NH phenotype. The newer culture media, 2i+LIF, was developed with the aim of eliminating the heterogeneity observed in Serum+LIF. An easy way to indirectly measure gene expression is by means of fluorescent proteins, tagging the gene of interest. More specifically a green fluorescent protein (GFP) can be added to tag Nanog, where the fluorescence is what is then measured by a special type of microscope which images florescence. The cell line commonly used to measure Nanog expression is known as TNGA, which was created by inserting a GFP cassette at the start codon (codons are groupings of three nucleotides which code for an amino acid) on one Nanog allele (an allele is a variant of a gene). TNGA mESCs show an almost complete elimination of the NL phenotype in 2i+LIF, suggesting that 2i+LIF enables complete, or 'ground-state', pluripotency *in-vitro* [159].

Recently, other cell lines have been developed, as it has been noticed that the degradation rates of Nanog and the GFP tagging Nanog in TNGA cells are different [39]. Due to the discrepancy, this fluorescent reporter is not a good proxy for observing Nanog dynamics. Therefore a newly developed cell line was created which uses a different fluorescent protein with comparable degradation rate to Nanog. The new cell line is known as ND [3, 4]. ND mESCs are still heterogeneous in Serum+LIF, although the difference between NH and NL is smaller. However, in 2i+LIF ND mESCs no longer show a complete elimination of the NL phenotype, with a small population of NL cells still present. This could be considered undesirable if the aim is to keep a population of pluripotent cells indefinitely.

Figure 2.2 shows concentrations of Nanog, measured by the fluorescent reporters, for the two cell line types in the different media. Figure 2.2a is for TNGA cells and shows the heterogeneity of Nanog in Serum+LIF (left panels of Figure 2.2a), which is lost in 2i+LIF (right panels of Figure 2.2a). Figure 2.2a also shows the same behaviour for Rex1, which is considered a direct target of Nanog, and measured as a proxy for it (see model in Section 2.2.3). Figure 2.2b is for the ND cell line which in Serum+LIF has a greater proportion of cells in a NL state than TNGA cells (Nd,

serum/LIF in Figure 2.2b). When the ND cells are grown in 2i+LIF the heterogeneity is reduced but still present, as shown by the small peak (Nd, 2i in Figure 2.2b, highlighted by *).

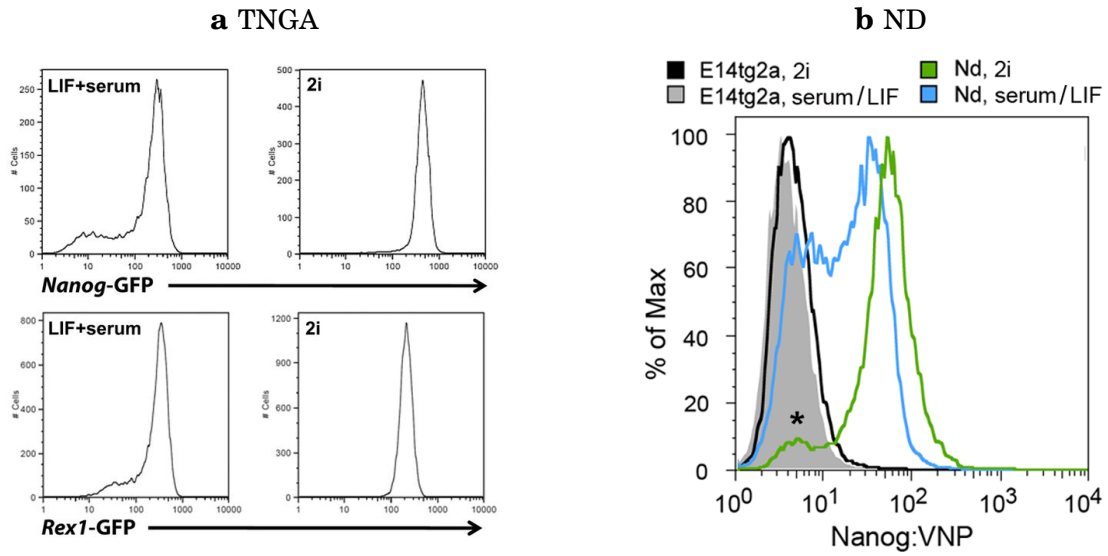


Figure 2.2: Concentrations of Nanog and its target Rex1 in different mESC line types. **a** TNGA cells, image adapted from [159]. **b** ND cells (VNP is venus nanog protein), image adapted from [3]. ND cells labeled Nd, with * highlighting persistence of NL population.

2.2 Modelling

2.2.1 Gene regulatory network (GRN) model

The central dogma of biology describes the process by which the information of a gene stored in DNA is transformed into a physical protein able to perform specific tasks. The process is DNA→mRNA→protein. The step from DNA to mRNA is called transcription and from mRNA to protein is called translation. Transcription is regulated by proteins called transcription factors (TF) which bind to certain regions of DNA, increasing or decreasing transcription. Proteins are the final output of (protein-encoding) genes, and can act as TFs which results in genes regulating other genes and even themselves. This forms a network of interacting genes, termed gene regulatory network (GRN) of genes activating and inhibiting other genes. Differential equations can be used to describe the changing protein levels, incorporating Hill equations to describe the interaction of a gene activating or inhibiting another gene. See Section 2.2.4 for more details on Hill equations and modelling GRNs.

2.2.2 Other mechanisms to model Nanog dynamics

Other modelling types have been put forward to explain the dynamic differences of pluripotent genes between the two media. A published review [86] outlined different ways of modelling Nanog dynamics, broadly characterized into four different modelling types: GRN (as above), probabilistic, allelic switching and protein-protein interaction networks.

GRNs describe transcriptional dynamics of interacting elements, which can involve positive and negative feedbacks and can also include stochastic elements. These models produce good agreement with experimental steady-state distributions and single cell dynamics.

Probabilistic models make no assumptions about underlying GRNs, thus they avoid over fitting of parameters and do not need to introduce assumptions. They are based on probability distribution functions and Gaussian noise. The example described in the review [78] had just one variable, that of Nanog.

An allelic switching model abstracted mESCs into four groups, each with a different state corresponding to altered Nanog behaviour [94]. Each group has a percentage of occupancy and there are transitional probabilities between the states described by a Markov chain. This model correctly recapitulated bistability in Serum+LIF, though was not used to model 2i+LIF.

A protein interaction network only considered protein-protein interactions, modelling transcription by a stochastic bursting mechanism [99]. One of their models correctly matched correlations between the genes Oct4 and Nanog (two of the genes considered part of the core pluripotency network, see Section 2.2.5), and matched their mRNA distributions. However it did make some incorrect predictions. This model was later improved by adding extra interactions making more predictions in line with experimental results.

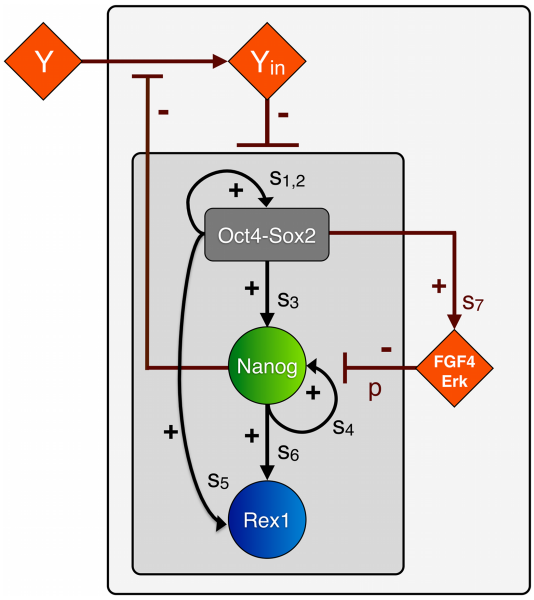
All these different models correctly match distributions of Nanog. However, none took into account the newer ND mESCs with their persistent NL cells. We therefore wanted to develop a model able to describe the persistence of NL cells in 2i+LIF, while accounting for factors present in different culture media.

2.2.3 Reference model

We looked at a previous model of a GRN [62] describing how the concentrations of Nanog and Rex1 differ between Serum+LIF and 2i+LIF. This model was fitted to experimental data generated using TNGA cells. Their GRN is shown in Figure 2.3a and was described mathematically by the stochastic differential equations (SDEs) in Figure 2.3b. This network is comprised of the genes Nanog, heterodimer Oct4-Sox2 and Rex1. Of note, Rex1 is only regulated by other factors,

and is therefore considered as an output to the model. Included into the model were signalling pathways of FGF4/Erk and Y/Yin (a differentiation signal we ignored as we were not interested in differentiation dynamics). Different media conditions were simulated via the value p, set at 15 for Serum+LIF and inhibited to 0 for 2i+LIF. Interactions between network elements were described mathematically with Hill equations, and noise added as a zero-mean Gaussian process. For more details about modelling GRNs, Hill equations and noise see Sections 2.2.4 and 2.2.7.

a



b

$$\frac{d[OS]}{dt} = \frac{s_{1,2}\cdot[OS]^2}{(1/k+[OS])^2\cdot d_O\cdot d_S} - d_{OS}\cdot(1+i_{OS}\cdot Y_{in})\cdot[OS] + \xi(0,\sigma_{OS})\cdot[OS]$$

$$\frac{d[N]}{dt} = \frac{s_3\cdot[OS]}{1/k+[OS]} + \frac{s_4\cdot[N]^2}{(1/k+[N]^2+p\cdot[E])} - d_N\cdot(1+i_N\cdot Y_{in})\cdot[N] + \xi(0,\sigma_N)\cdot[N]$$

$$\frac{d[R]}{dt} = \frac{s_5\cdot[OS]}{1/k+[OS]} + \frac{s_6\cdot[N]^2}{1/k+[N]^2} - d_R\cdot(1+i_R\cdot Y_{in})\cdot[R] + \xi(0,\sigma_R)\cdot[R]$$

$$\frac{d[E]}{dt} = \frac{s_7\cdot[OS]}{1/k+[OS]} - d_E\cdot[E]$$

Figure 2.3: Reference GRN model of [62]. **a** Diagram of Network interactions. Core network is within the inner grey box, composing Nanog, heterodimer Oct4-Sox and Rex1. Respective transcription rates are denoted by S_i. Additional signalling is included; FGF4/Erk inhibiting Nanog with rate p and a differentiation signal Y (which we ignore in this thesis). **b** Equations describing the rates of genes involved in the network using Hill terms, resulting from the GRN interactions. S_i are maximum transcription rates, d_i are degredation rates, $1/k$ is the Michaelis–Menten constant and σ_i are noise amplitudes added to each network factor.

The developers of the reference GRN model simulated it 10,000 times, and recorded the concentration of Rex1 at the end of each run. The resulting distribution of Rex1 values are shown in Figure 2.4a and 2.4b for Serum+LIF and 2i+LIF respectively, on top of measured Rex1 data. Their model correctly matched TNGA Rex1 tagged data, with a complete elimination of NL cells in 2i+LIF. However it has not been fitted to distributions observed in ND cells.

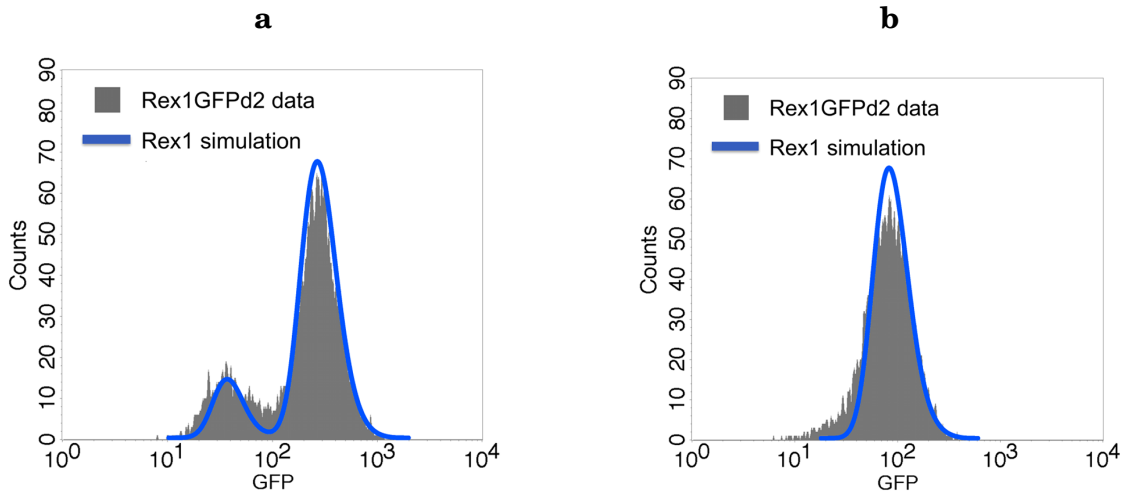


Figure 2.4: Simulation results of Rex1 distributions in reference model compared to Rex1 data. Images adapted from [62]. **a** Serum+LIF results. **b** 2i+LIF results.

2.2.4 How to model a GRN

The reference model used a GRN to describe the changing behaviour of Nanog between the two media. Below we derive Hill equations showing how they model transcription saturations, and we demonstrate how to incorporate them into the mathematical framework of a GRN.

Hill equation derivation

The derivation described here is based on the work in [6]. Transcription is the process by which mRNA is produced in a cell requiring a special protein called a transcription factor (TF) to bind to a certain region of the DNA. TFs either activate the transcription process or they inhibit it, and as the concentration of TF increases the rate at which transcription happens saturates. Hill equations are commonly used to describe transcription rate saturation in GRNs. A Hill term for activation is

$$\frac{S \cdot [A]^n}{K^n + [A]^n}, \tag{2.1}$$

and for inhibition is

$$\frac{S \cdot K^n}{K^n + [A]^n}. \tag{2.2}$$

Concentrations are denoted in square brackets ([ ]). S is the maximum rate, $[A]$ is the concentration of substrate of interest, n is the Hill coefficient describing cooperativity and K is the

Michaelis–Menten constant which is also the concentration of substrate ($[A]$) required for the Hill term to be half of its maximum value.

Deriving the Hill equation starts by considering the process of TF binding. Figure 2.5 shows an illustration of a TF (A in diagram) and a binding site on the DNA (B in diagram) to which the TF can attach. When the TF is attached to the binding site it forms a complex (C in diagram) and transcription can now happen.

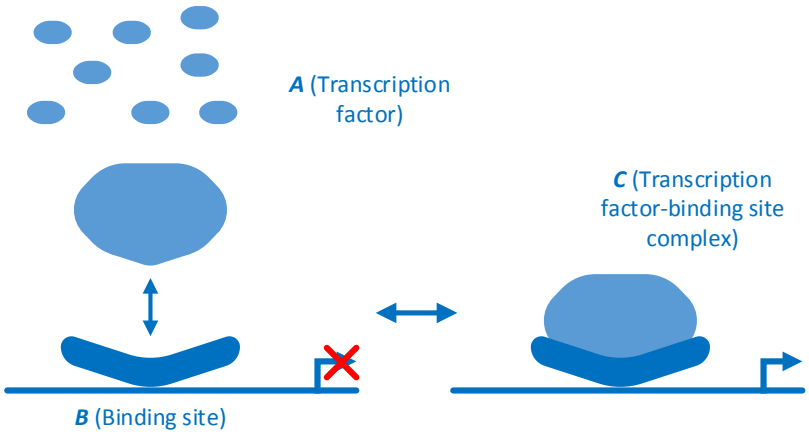


Figure 2.5: Illustration of a transcription factor (A) attaching to a binding site (B), forming a binding site complex (C), allowing transcription (indicated by bent arrow) to occur.

This process can be described in a reversible rate equation given by

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C, \tag{2.3}$$

where A is the transcription factor, B is the binding site, C is the factor-complex formed, k_1 is the rate at which the complex is formed and k_2 is the rate at which the complex is broken back into the TF and free binding site.

The law of mass action states that the rate of a reaction is linearly proportional to the concentration of its reactants, and can be used to define differential equations describing how the concentrations of the three reactants (TF, binding site and complex) change over time. The differential equations for the reactants are given by

$$\frac{d[A]}{dt} = k_2[C] - k_1[A][B], \tag{2.4}$$

$$\frac{d[B]}{dt} = k_2[C] - k_1[A][B], \tag{2.5}$$

and

$$\frac{d[C]}{dt} = k_1[A][B] - k_2[C] \tag{2.6}$$

for TF, binding site and complex, respectively.

These differential equations can be simplified by assuming that all the reactants are in a 'quasi steady-state'. This assumes that the process by which the TF binds and unbinds happens so quickly that in a sense it has already happened and therefore the rate of changes can be set to zero, given by

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = \frac{d[C]}{dt} = 0 \implies [C] = \frac{k_1}{k_2}[A][B]. \tag{2.7}$$

The rate of mRNA production is proportional to the portion of binding sites that are bound by a TF, with added degradation. The differential equation describing the production of mRNA is given by

$$\frac{d[m]}{dt} = S \cdot \frac{[C]}{[B]+[C]} - d_m \cdot [m], \tag{2.8}$$

where m is mRNA concentration, S is a proportionality constant equivalent to the maximum transcription rate and d_m is the degradation rate. Substituting in the terms found from quasi steady-state (equation (2.7)) and taking out common factors leads to

$$\frac{d[m]}{dt} = S \cdot \frac{\frac{k_1}{k_2}[A][B]}{[B] + \frac{k_1}{k_2}[A][B]} - d_m \cdot [m] = S \cdot \frac{[B]\frac{k_1}{k_2}}{[B]\frac{k_1}{k_2}} \cdot \frac{[A]}{\frac{k_2}{k_1} + [A]} - d_m \cdot [m]. \tag{2.9}$$

The final Hill equation describing how the rate of mRNA production is affected by a TF, for the case when the TF acts as an activator is shown, with added degradation by

$$\frac{d[m]}{dt} = \frac{S \cdot [A]}{K + [A]} - d_m \cdot [m], \quad K = \frac{k_2}{k_1}. \tag{2.10}$$

The case when the TF acts as an inhibitor is represented in Figure 2.6. This assumes that the gene is already in the process of transcribing before the TF binds. After binding, transcription is stopped.

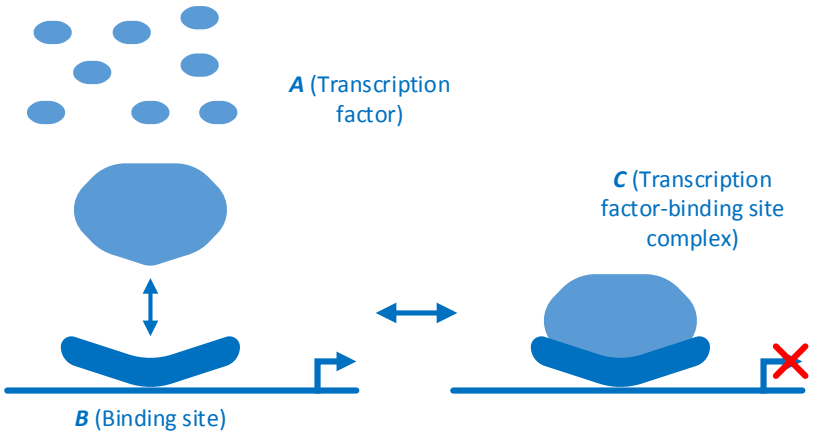


Figure 2.6: Illustration of a transcription factor (A) attaching to a binding site (B), forming a binding site complex (C), stopping transcription (red cross).

The case for inhibition is the same as for activation as in equations (2.3) to (2.7), except the rate of mRNA production is now proportional to the portion of binding sites not bound by a TF minus some degradation, as shown by

$$\frac{d[m]}{dt} = S \cdot \frac{[B]}{[B]+[C]} - d_m \cdot [m]. \tag{2.11}$$

Substituting in the terms found from quasi steady-state (equation (2.7)) and taking out common factors leads to

$$\frac{d[m]}{dt} = S \cdot \frac{[B]}{[B]+\frac{k_1}{k_2}[A][B]} - d_m \cdot [m] = S \cdot \frac{[B]\frac{k_1}{k_2}}{[B]\frac{k_1}{k_2}} \cdot \frac{\frac{k_2}{k1}}{\frac{k_2}{k_1}+[A]} - d_m \cdot [m]. \tag{2.12}$$

The final Hill equation describing how the rate of mRNA production is affected by a TF, for the case when the TF acts as an inhibitor and with added degradation, is given by

$$\frac{d[m]}{dt} = \frac{S \cdot K}{K+[A]} - d_m \cdot [m], \quad K = \frac{k_2}{k_1}. \tag{2.13}$$

We examine one more scenario, where two TFs form a homodimer and activate transcription together as illustrated in Figure 2.7.

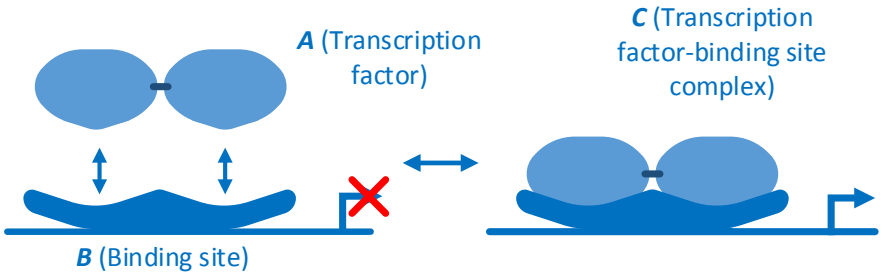


Figure 2.7: Illustration of two transcription factors (A) formed as a homodimer, attaching to a binding site (B), forming a binding site complex (C), allowing transcription (indicated by bent arrow) to occur.

This process can be described by a reversible rate equation given by

$$2A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C. \tag{2.14}$$

Here the reaction requires two TFs, where A is a transcription factor, B is the binding site, C is the factor-complex formed, k_1 is the rate at which the complex is formed and k_2 is the rate at which the complex is broken back into TFs and free binding site. The law of mass action is again used to describe the rate of change of concentrations of the three reactants (TF, binding site and complex) over time and is given by

$$\frac{d[A]}{dt} = k_2[C] - k_1[B][A]^2, \tag{2.15}$$

$$\frac{d[B]}{dt} = k_2[C] - k_1[B][A]^2 \tag{2.16}$$

and

$$\frac{d[C]}{dt} = k_1[B][A]^2 - k_2[C] \tag{2.17}$$

for TF, binding site and complex, respectively.

Assuming again that all the reactants are in a quasi steady-state leads to

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = \frac{d[C]}{dt} = 0 \implies [C] = \frac{k_1}{k_2}[B][A]^2. \tag{2.18}$$

The rate of mRNA production is again proportional to the portion of binding sites bound by TFs, with added degradation. The differential equation describing the production of mRNA for the TF homodimer activating transcription is given by

$$\frac{d[m]}{dt} = S \cdot \frac{[C]}{[B]+[C]} - d_m \cdot [m], \tag{2.19}$$

where m is mRNA concentration, S is the proportionality constant and d_m is the degradation rate. Substituting in equation (2.18) and simplifying is shown by

$$\frac{d[m]}{dt} = S \cdot \frac{\frac{k_1}{k_2}[B][A]^2}{[B]+\frac{k_1}{k_2}[B][A]^2} - d_m \cdot [m] = S \cdot \frac{[B]\frac{k_1}{k_2}}{[B]\frac{k_1}{k_2}} \cdot \frac{[A]^2}{\frac{k_2}{k_1}+[A]^2} - d_m \cdot [m]. \tag{2.20}$$

The final Hill equation describing how the rate of mRNA production is affected by two TFs, for the case when they act cooperatively as activators and with added degradation, is given by

$$\frac{d[m]}{dt} = \frac{S \cdot [A]^2}{K^2+[A]^2} - d_m \cdot [m], \quad K = \sqrt{\frac{k_2}{k_1}}. \tag{2.21}$$

For two TFs cooperatively inhibiting transcription, mRNA rate is proportional to portion of binding sites not bound by a TF homodimer, and is given by

$$\frac{d[m]}{dt} = \frac{S \cdot K^2}{K^2+[A]^2} - d_m \cdot [m], \quad K = \sqrt{\frac{k_2}{k_1}}. \tag{2.22}$$

Extending this to even more TFs acting cooperatively is simply a case of increasing the exponents of A and K. This exponent is known as the Hill coefficient. The rate of mRNA production for a general number of TFs, n, cooperatively activating is given by

$$\frac{d[m]}{dt} = \frac{S \cdot [A]^n}{K^n+[A]^n} - d_m \cdot [m], \quad K = \sqrt[n]{\frac{k_2}{k_1}}, \tag{2.23}$$

and for cooperatively inhibiting is given by

$$\frac{d[m]}{dt} = \frac{S \cdot K^n}{K^n+[A]^n} - d_m \cdot [m], \quad K = \sqrt[n]{\frac{k_2}{k_1}}. \tag{2.24}$$

This only describes how the concentration of mRNA is affected by a TF. Describing how the end protein concentration is affected requires one more step. Transcription dynamics can be assumed to occur on a much faster time scale than translation, and the quasi steady-state assumption can be made, setting the rate of change of mRNA to zero given by

$$\frac{d[m]}{dt} = \frac{S \cdot [A]^n}{K^n + [A]^n} - d_m \cdot [m] = 0, \tag{2.25}$$

leading to

$$[m] = \frac{1}{d_m} \cdot \frac{S \cdot [A]^n}{K^n + [A]^n}. \tag{2.26}$$

The rate of change of protein is assumed to be proportional to the concentration of mRNA, minus protein degradation as given by

$$\frac{d[p]}{dt} = \alpha[m] - d_p[p], \tag{2.27}$$

where α is a proportionality constant and d_p is the degradation rate of the protein. This form ignores basal/background transcription that can occur due to random movements of the DNA strands and other unexplained molecular mechanisms.

The final equation describing how the concentration of a gene's protein (p) is regulated by a TF (A) is found by substituting equation (2.26) into (2.27) and is given by

$$\frac{d[p]}{dt} = \frac{S_p \cdot [A]^n}{K^n + [A]^n} - d_p[p], \quad S_p = \frac{\alpha \cdot S}{d_m} \tag{2.28}$$

and

$$\frac{d[p]}{dt} = \frac{S_p \cdot K^n}{K^n + [A]^n} - d_p[p], \quad S_p = \frac{\alpha \cdot S}{d_m} \tag{2.29}$$

for activation and inhibition, respectively. S_p is the maximum protein production rate and K is the Michaelis–Menten constant.

Describing a GRN with Hill equations

Hill equations can be used to describe the interactions of a GRN under a few assumptions. The process of transcription in an individual cell is binary, it is either happening or it is not. The Hill equation is a smooth function ranging between transcription and no transcription. Therefore it can only be applied when considering a large of number of cells, modelling a distribution of transcription in a population.

How to apply Hill equations to describe a GRN is described here by an example. Figure 2.8 shows two simple made up gene networks, with Figure 2.8a being a single gene, B, activated

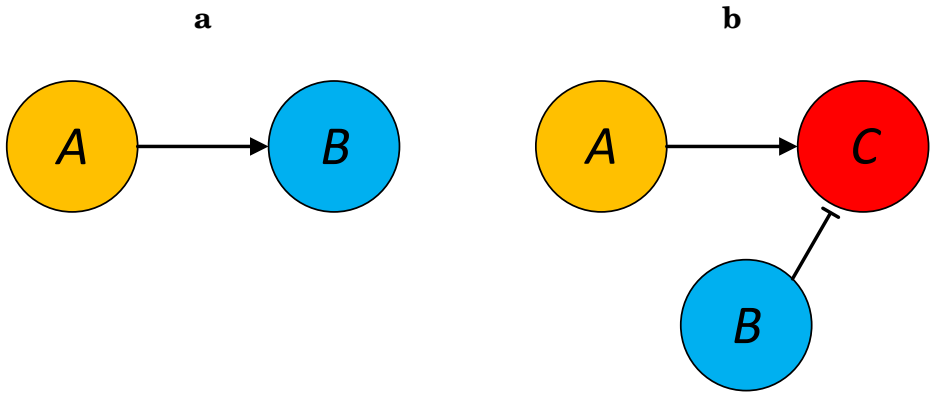


Figure 2.8: Two simple GRNs to demonstrate application of Hill equations to describe gene interactions. **a** Simple gene network with single activation (A activating B indicated by arrow). **b** Simple gene network with single activation and single inhibition (A activating C indicated by arrow, and B inhibiting C indicated by flat-headed arrow).

by another, A. Figure 2.8b shows a single gene, C, being regulated by two other genes; A which activates it, and B which inhibits it. This example is to demonstrate how to describe a gene's affect on another gene, therefore only Gene B in Figure 2.8a and C in Figure 2.8b are considered.

The activation of A on B in Figure 2.8a can be described with a Hill equation as given by

$$\frac{d[B]}{dt} = \frac{S_A \cdot [A]^n}{K^n + [A]^n} - d_B[B]. \tag{2.30}$$

The regulation of A and B on C, as in Figure 2.8b, can be described as given by

$$\frac{d[C]}{dt} = \frac{S_A \cdot [A]^n}{K^n + [A]^n} + \frac{S_B \cdot K^n}{K^n + [B]^n} - d_C[C], \tag{2.31}$$

where two Hill terms (one for activation, one for inhibition) are added together. This assumes the regulation from TFs act as an OR gate. The Hill coefficient n, as well as the other parameters (K, the Michaelis–Menten parameter; d_i and S_i, the degradation rate and maximum transcription rate respectively of substrate i) can be chosen based on experimental data or desired system behaviour. This method can then be extended to a larger GRN, forming differential equations for each element of the network by summing Hill terms associated to every element which acts on it.

2.2.5 Bistability in an idealised 'core' GRN

Nanog, Oct4 and Sox2 are considered the core genes involved in pluripotency and cell fate in mESCs. The behaviour of Nanog has been shown to be bistable in mESCs in Serum+LIF and 2i+LIF (bistable in 2i+LIF for ND cells not TNGA) [3, 4, 159], displaying two different stable

levels of the gene, termed Nanog high (NH) and Nanog low (NL). The genes Oct4 and Sox2 act together as a heterodimer [23] and its concentration levels have been shown to be stable at a single concentration level, with little fluctuation [19, 138]. We recapitulate how the bistability of Nanog can be derived from a simple ODE in this section, starting from the reference GRN and simplifying it.

Figure 2.9 shows two different GRN interactions of the core genes, suggested in [23] for Figure 2.9a and [62] for Figure 2.9b (we removed Rex1 as it is not considered a part of the core network). Oct4-Sox2 is a heterodimer, whose process of formation and disintegration is shown in the GRN of Figure 2.9a. Figure 2.9b simplifies the Oct4-Sox2 heterodimer interaction to a simple auto-activation. The network in Figure 2.9b also removed the activation of Oct4 and Sox2 by Nanog as in Figure 2.9a to simplify the network. They were not needed to explain the main bistable behaviour of Nanog and add extra complexity and parameters to fit. We took the simplified core network of Figure 2.9b and simplified it further to demonstrate how the bistability of Nanog could arise from a positive feedback loop on Nanog.

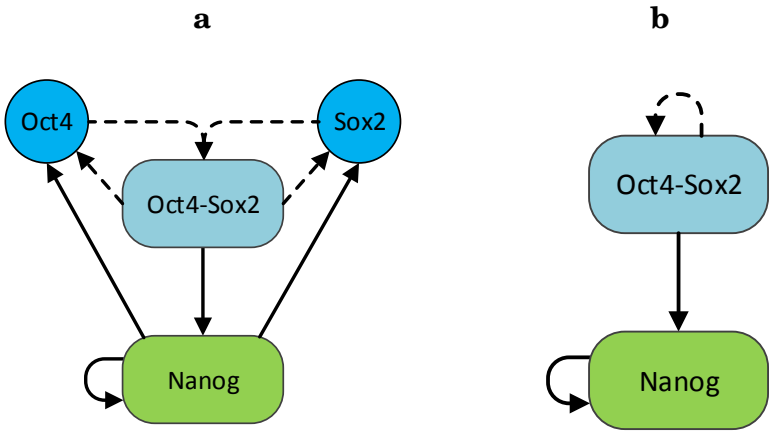


Figure 2.9: Two different GRN schematics of the core genes involved in pluripotency in mESCs. Solid lines are transcriptional interactions, dashed are non-transcriptional. **a** Core pluripotency network as described by [23]. This shows interactions between the genes Oct4, Sox2, their heterodimer Oct4-Sox2 and Nanog.**b** Simplified network as in [62] showing only interaction between the Nanog and the heterodimer Oct4-Sox2. Activation of Oct4 and Sox2 by Nanog removed as these interactions were not necessary to describe Nanog dynamic.

The final simplification we made was assuming that the activation of Oct4-Sox2 on Nanog is weak in comparison to Nanog's activation of itself. As the distributions of Oct4-Sox2 concentrations are also monostable, Oct4-Sox2 activation on Nanog can be assumed to be constant. This simplifies the underlying equations of the system to a single ODE which can be easily analysed,

given by

$$\frac{d[N]}{dt} = S_{OS} + \frac{S_N \cdot [N]^2}{K^2 + [N]^2} - d_N \cdot [N], \tag{2.32}$$

where S_{OS} is a small constant representing the weak activation of Nanog by Oct4-Sox2.

This focuses the analysis on the key changing gene Nanog only. Nanog has been shown to form homodimers [98, 153], motivating a choice of Hill coefficient of 2 for this basic system. We analysed this system graphically by splitting the RHS into two parts, $f(N)$ and $g(N)$, given by

$$\frac{d[N]}{dt} = f(N) - g(N). \tag{2.33}$$

The two parts $f(N)$ and $g(N)$ are defined as

$$\begin{cases} f(N) = S_{OS} + \frac{S_N \cdot [N]^2}{K^2 + [N]^2} \\ g(N) = d_N \cdot [N] \end{cases} \tag{2.34}$$

where $f(N)$ is the positive contributions to the rate of change and $g(N)$ the negative parts. Steady-state points are defined as when the rate of change is zero, as in

$$\frac{d[N]}{dt} = 0. \tag{2.35}$$

Substituting equation (2.35) into (2.33) steady-state points are found by the simple equality given by

$$f(N) = g(N). \tag{2.36}$$

We find steady-state points by drawing $f(N)$ and $g(N)$ and seeing where the curves intersect. Figure 2.10 shows a plot of the two functions and their three intersections, or steady-states. It should be noted that these are not nullclines, they are graphical representations of the functions $f(N)$ and $g(N)$ as stated in equation (2.34). The final consideration in this system is the stability of the steady-states. This is done by referring back to equation (2.33), where the rate of change is positive for $f(N) > g(N)$ and negative for $g(N) > f(N)$. The lower part of Figure 2.10 shows the 1-dimensional space of the concentration [N]. Arrows indicate the direction of changing concentration, when $f(N) > g(N)$ the rate of change of concentration is positive and so points to the right. When $g(N) > f(N)$ it is to the left. Steady-states where the arrows point into them are stable, and where they point away from are unstable. From this diagram it can be seen that two

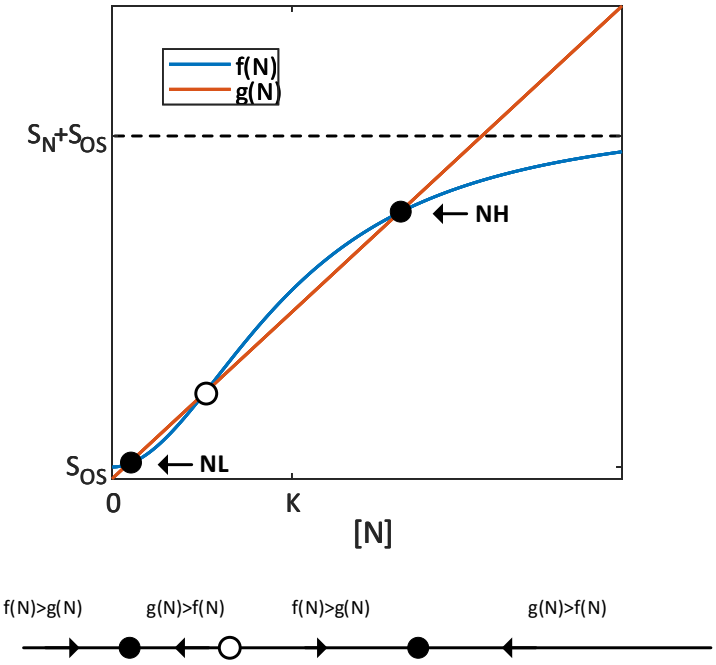


Figure 2.10:
Graphical representation of steady-states for the differential equation: $\frac{d[N]}{dt} = f(N) - g(N)$ where $f(N)$ is a Hill equation and $g(N)$ is linear degradation, see equation (2.34). Intersections of the two curves are where the rate of change of the system is zero. Stability of these states are determined by looking at the system flow directions either side of them. Stable steady-states are indicated with ● and unstable steady-states with ○.

of the steady-states are stable and the middle one is unstable.

This GRN therefore exhibits bistable behaviour for Nanog, under certain parameter conditions; one NL and one NH stable steady-state separated by an unstable steady-state between them. The bistability is due to the non linear positive feedback loop on Nanog, and Nanog's degradation.

2.2.6 Parameter effects on bistability in the idealised 'core' GRN

There are four parameters in this system; S_N, S_{OS}, K and d_N. The values of these parameters change the stability behaviour of the core GRN. How these parameters affect the system can be explored graphically. The degradation rate is kept constant throughout the following exploration, as seen by the linear diagonal line in all the Figures. Maximum transcription rate is indicated by a horizontal dashed black line on all the Figures. These demonstrations are qualitative and so no

specific values are given.

The system stability when changing parameter S_{OS} can be seen in Figure 2.11. Here the value is increasing from left to right which effectively transposes the Hill function higher. As the value increases the lower two intersections of the curve come closer together (Figure 2.11a) and eventually the curves pull apart at a Saddle-node bifurcation (Figure 2.11b), leaving only a single high stable steady-state (Figure 2.11c).

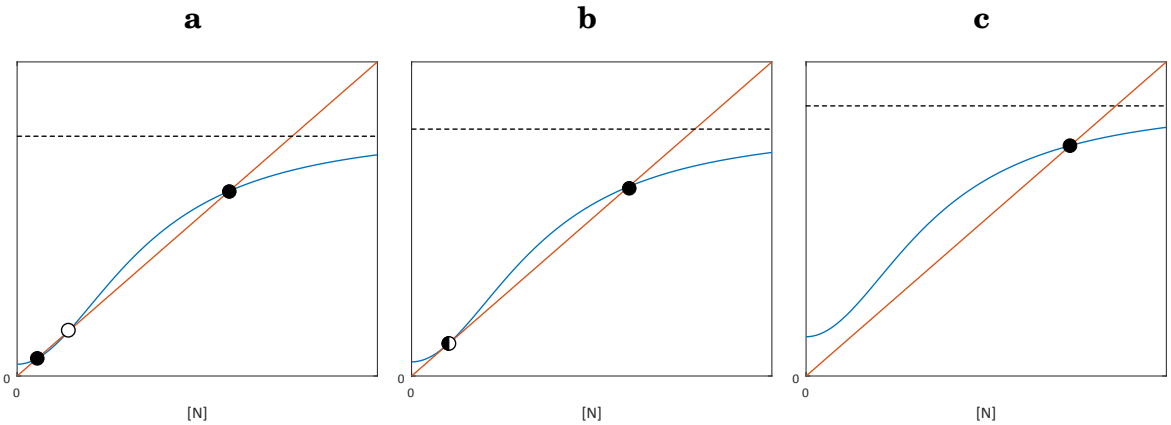


Figure 2.11: Graphs demonstrating change in bistability as S_{OS} value increases **a** to **c**. Black dots ● represent stable steady-states, black-outlined white dots ○ represent unstable states and half-filled dots ◐ represent Saddle-node bifurcations.

The system stability when changing parameter S_N can be seen in Figure 2.12. Here the value is increasing from left to right, effectively increasing the maximum gene production rate. As the value increases the non linearity of the Hill term bends it away from the degradation curve (Figure 2.12a), eventually bending away enough to form a Saddle-node bifurcation (Figure 2.12b) and then bending fully away leaving only a single high stable steady-state (Figure 2.12c).

The system stability when changing parameter K can be seen in Figure 2.13. Here the value is decreasing from left to right, increasing the systems sensitivity to increasing TF concentrations. As the value decreases the Hill term curves more away from the degradation curve (Figure 2.13a), eventually forming a Saddle-node bifurcation (Figure 2.13b) and then leaving only a single high stable steady-state (Figure 2.13c). Note the maximum rate (dashed black line) remains constant throughout changing K.

Another parameter to consider is the Hill coefficient which has previously been set to 2 for all the examples. The equations of the graphs being considered, this time with variable Hill coefficient n, are

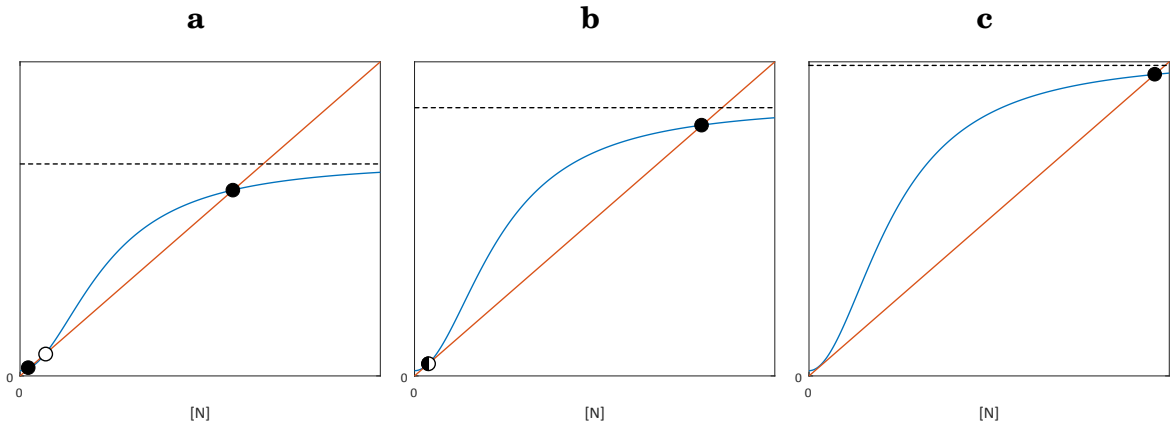


Figure 2.12: Graphs demonstrating change in bistability as S_N value increases **a** to **c**. Black dots ● represent stable steady-states, black-outlined white dots ○ represent unstable states and half-filled dots ◐ represent Saddle-node bifurcations.

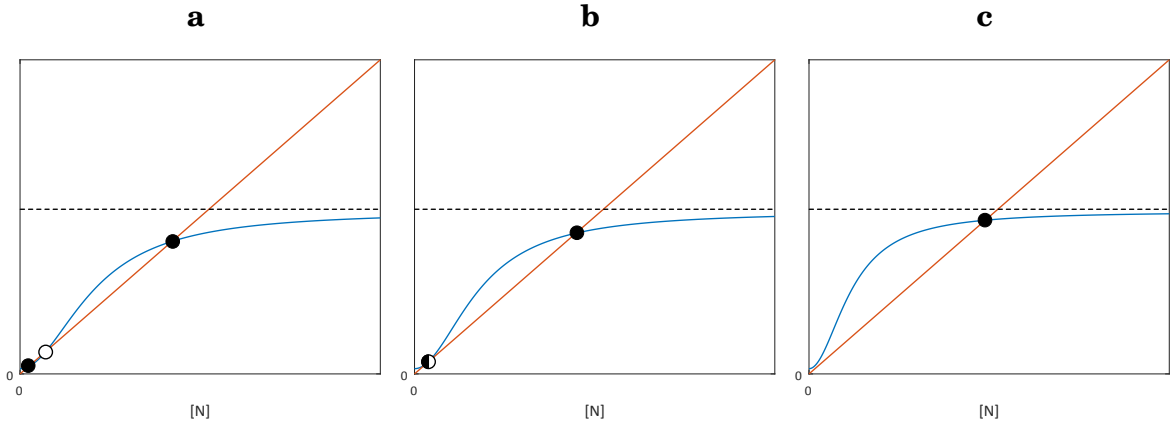


Figure 2.13: Graphs demonstrating change in bistability as the Michaelis–Menten paramater K decreases **a** to **c**. Black dots ● represent stable steady-states, black-outlined white dots ○ represent unstable states and half-filled dots ◐ represent Saddle-node bifurcations.

$$\begin{cases} f(N) = S_{OS} + \frac{S_N \cdot [N]^n}{K^n + [N]^n} \\ g(N) = d_N[N]. \end{cases} \tag{2.37}$$

Figure 2.14 shows the system as the Hill coefficient is increased, Figure 2.14a to Figure 2.14c. This coefficient is the main source of non linearity in the system, which displays more non linear behaviour as the value is increased.

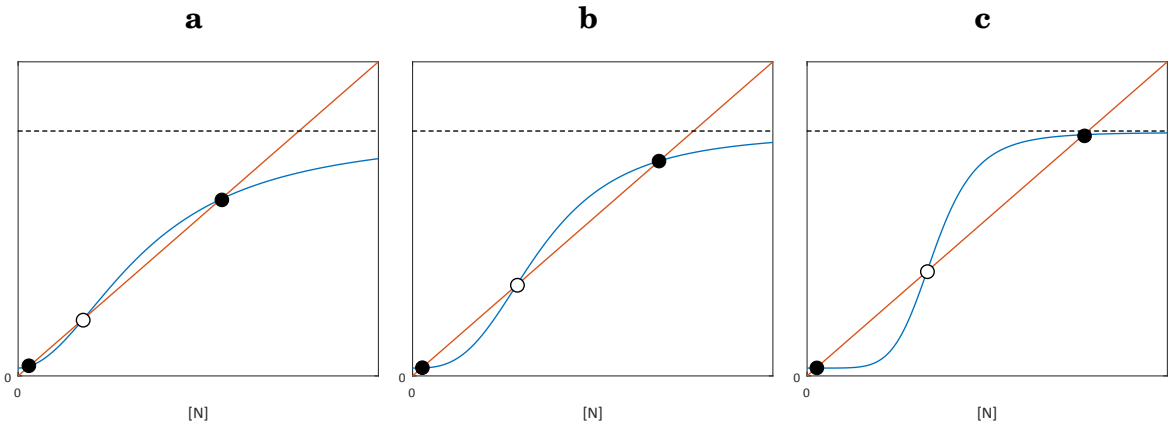


Figure 2.14: Graphs demonstrating change of system graphs as Hill coefficient, n, is increased **a** to **c**. Black dots ● represent stable steady-states and black-outlined white dots ○ represent unstable states.

The previous principles can be used to alter the stability of the deterministic system to a desired level. For example if the wanted behaviour is for the steady-states to be brought closer together as in Figure 2.15. This is achieved here by simultaneously decreasing S_n and K.

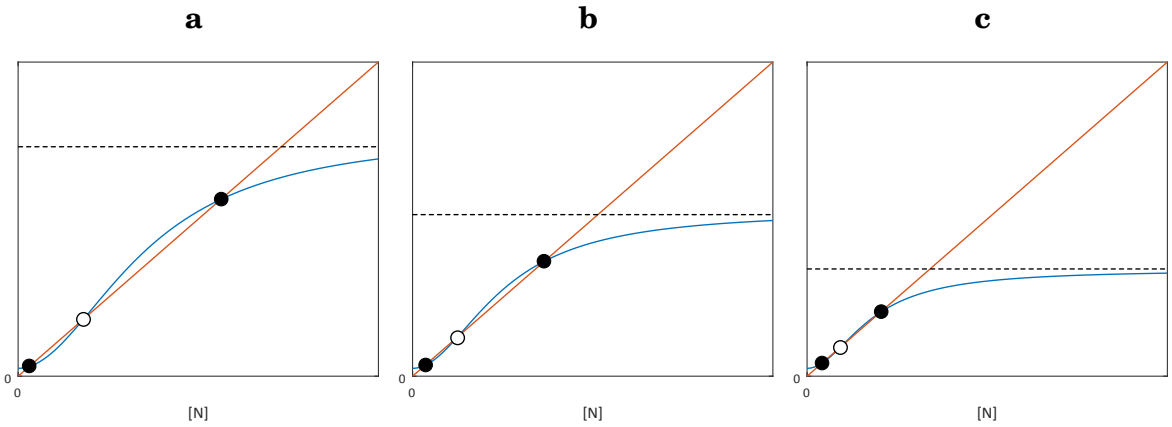


Figure 2.15: Graphs demonstrating change in bistability as S_n and k values simultaneously decrease **a** to **c**. Black dots ● represent stable steady-states and black-outlined white dots ○ represent unstable states.

2.2.7 Stochasticity to encompass complex molecular processes

The system shown so far is deterministic, and any simulation of a cell encompassing that GRN will always display the same concentration of Nanog, either the NH or the NL steady-state with no intermediate values. Real mESCs display a continuous distribution of concentration values due to the fact that they have many more genes, with very complicated regulations and molecular processes. With current knowledge and technologies, modelling the exact regulations

of the huge number of processes in a cell is impossible. Therefore noise is added to the system to approximate it. This changes the ODE from Section 2.2.5 (equation (2.32)) into a stochastic differential equation (SDE) of the form given by

$$\frac{d[N]}{dt} = S_{OS} + \frac{S_N \cdot [N]^2}{K^2 + [N]^2} - d_N \cdot [N] + \zeta(0, \sigma_N) \cdot [N]. \tag{2.38}$$

The noise is applied as a random variable ζ drawn from a normal distribution with mean 0 and standard deviation σ_N, multiplied by the concentration of the gene, i.e. $[N]$, and added to the RHS as in [62]. Simulating a large number of cells produces a distribution of concentration values for Nanog. The two main ways of adding noise are multiplicative, as described here, and additive where the noise term is added without multiplying by the gene's concentration. We chose to follow the multiplicative approach as in the reference model, due to it better describing the greater variability in NH as compared to NL; i.e. NH cells are noisier than NL and additive noise would not capture this behaviour.

2.2.8 Interplay between noise strength 'σ_N' and maximum transcription rate of Nanog 'S_N'

The GRN of the reference model in Section 2.2.3 demonstrates how, in their GRN, simultaneously changing the maximum transcription rate of Nanog, S_4 equivalent to S_N in equation (2.38), and the noise strength can affect the proportion of cells in NH and NL. Increasing Nanog's auto-regulatory rate decreases the proportion of NL cells, whilst increasing Nanog's transcriptional noise increases the proportion of NL cells.

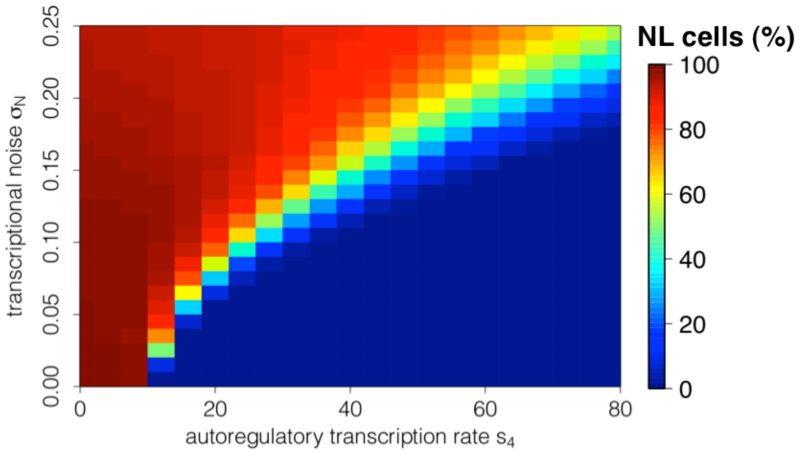


Figure 2.16: Diagram showing how the proportion of NL (and inversely NH) cells is altered when changing transcriptional noise and Nanog's auto-regulatory rate, adapted from [62].

We assume this behaviour qualitatively holds for any GRN of Nanog involving positive feedback and multiplicative noise.

2.3 Extending the reference model

The model we developed starts with the reference model of Section 2.2.3, adding more genes and molecular interactions to better explain the dynamics underlying pluripotency in mESCs. The reference model only included one of the two inhibitors present in 2i+LIF, PD. Therefore we wanted the new network to describe Chiron as well. We additionally wanted the new network to incorporate genes related to cell-cycle and pluripotency as these have been shown to be affected in the two media. Additionally the reference model was fitted to TNGA cell data, therefore the new model was fitted to measured ND cell distributions. The work flow to add these extra genes and interactions is summarized in Figure 2.17; below we explain the types of experiment used, how additional genes and their interactions were inserted, and how the inputs present in the different culture media were accounted for.

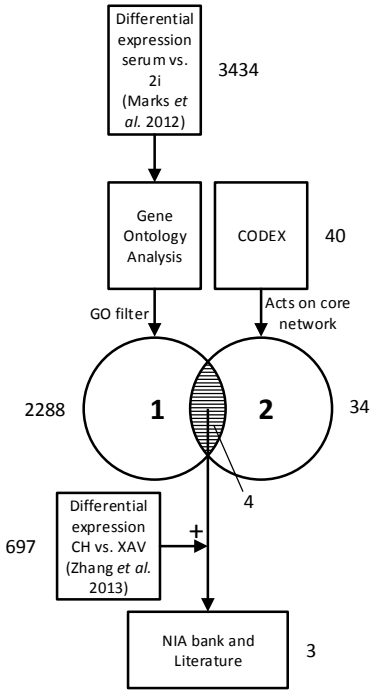


Figure 2.17: Schematic of the process for identifying transcriptional interactions not present in reference model. Set '1' is the output from gene ontology analysis, and set '2' is the output from the CODEX analysis. Numbers next to each item indicate the number of genes in the respective step.

2.3.1 Types of experimental data used

We used a number of data sources to expand the GRN, each one containing results from performing different types of experiment. Three different experiment types were used by the sources; Rna-seq, ChiP-seq and microarray. Descriptions of these different experiments are in Table 2.1. Citations for the actual data sources used will be given in their relevant section.

Type of Experiment	**Description**
Rna-seq	Transcribed mRNA from cells are extracted and fragmented. Each fragment is reverse transcribed into corresponding fragments of nucleotides called complementary DNA (cDNA). The fragments of cDNA are then sequenced and aligned to a reference genome sequence to reconstruct which genome regions were transcribed. Counting the number of fragments for a specific part of the genome gives an expression level for that part.
Chip-Seq	Chromatin immunoprecipitation sequencing. Protein of interest is crosslinked (chemically bounded) to DNA of cells *in-vivo* with formaldehyde, and the DNA of the cell is then isolated. DNA is then broken into segments of around 500 base pairs using sonication. Anti-body specific to the protein of interest is added to isolate fragments of DNA bounded by the protein, and to reverse the cross linking, leaving only fragments of DNA previously bound by the protein. The DNA is copied using polymerase chain reaction resulting in a sample of DNA bound by the protein. The samples are then sequenced and mapped to the genome it came from to identify where the protein bound. This method can only reveal whether a TF has bound to the regulatory region of a gene, i.e. it cannot say whether the TF activates or inhibits the transcription.
Microarray	A series of microscopic pits, where each pit contains specially engineered strands of DNA. Each pit produces a different target gene, which in turn produces a fluorescent protein. A TF is added to each pit and the change in fluorescence measured, with increasing fluorescence corresponding to the TF increasing the target gene's expression and vice versa for a decrease.

Table 2.1: Table showing different types of experiments used to expand the reference GRN.

2.3.2 Additional genes

We re-analysed a published data set of genes differentially expressed between the two media [83], genes expressed higher in Serum+LIF vs 2i+LIF and vice versa. This data set used Rna-seq to compare expression levels of genes between the two media.

We performed Gene Ontology (GO) analysis using the platform DAVID [64], an online tool for GO classification. GO is the labelling of genes by their associated biological process (BP). For each BP, DAVID looks at the number of genes in an input list that are associated with that BP. It also looks at the number of genes in the entire organism also associated with that BP. DAVID then calculates the probability that a random sample from the entire genome (with equal number of genes from the input list) has an equal or greater proportion of genes associated with the BP. The BPs are sorted in order of significance (lowest highest p-value) along with their associated genes. The genes differentially expressed between the two media, listed in [83], was put into DAVID. Only genes whose BPs related to development and differentiation, proliferation, cell-cycle, morphology or cell death were retained. In total, 3434 genes were input into DAVID which output 2288 related to the BPs of interest (set '1' in Figure 2.17).

We selected only genes which act on the core network (i.e., Oct4, Sox2 and Nanog) because only these would be capable of affecting the dynamics of Nanog, and therefore indirectly pluripotency. The published data set does not have information on how the genes in it regulate other genes, therefore a different source is needed to add this information. An updated compendium of mESC TFs was used called CODEX [85, 127] which contains what genes 40 mESC TFs regulate. CODEX contained ChiP-seq experiment results to supply this information. From CODEX, a list of 34 genes (set '2' in Figure 2.17) was generated, where each of these genes regulates at least one of the core factors.

Genes which were both found to be involved in the BPs of interest and also to regulate one of the core factors (intersection of set '1' and set '2' in Figure 2.17) were considered of interest to add to the reference model. We identified four genes; Mycn, Rest and Chd7 expressed higher in Serum+LIF; and Prdm14 expressed higher in 2i+LIF. Mycn, Rest and Prdm14 were found in previous literature to significantly affect mESC pluripotency [140, 150, 164]. However Chd7 was found to not be essential for self-renewal and pluripotency [130] and was therefore not included into the final network.

2.3.3 Gene interaction directions

The three genes found (Mycn, Rest and Prdm14) are known to regulate one of the core factors (Oct4, Sox2 and Nanog). However, the analysis so far cannot say whether the interactions involve activations or inhibitions. For this, we used the NIA bank [107], which contains results from

microarray experiments with 54 different TFs. Out of the six genes in the extended network, four were in NIAbank: Nanog, Oct4, Sox2 and Mycn.

2.3.4 Medium signalling components and protein-protein interactions

The culture 2i+LIF contains two inhibitors; PD and Chiron. PD is, as in [62], modelled as a signal which weakens the inhibitory effect of FGF4/Erk signalling on Nanog. However the reference model ignored the effects of Chiron. Chiron is an inhibitor of GSK3 [7], which in turn destabilizes β-catenin [59, 161] increasing its degradation (Chiron inhibits GSK3 which inhibits β-catenin). In effect therefore, Chiron acts as an indirect activator of β-catenin. Two other important non-transcriptional interactions of this pathway are β-catenin inhibiting Tcf3 and Tcf3 inhibiting the regulation of Oct4-Sox2 on Nanog [158]. Given experimental evidence that PD alone is unable to sustain mESC clonal propagation [57, 158] and the key role of the β-catenin pathway in pluripotency and reprogramming of somatic cells [72, 87], we explored these additional interactions.

2.3.5 Putting it all together

Table 2.2 shows which genes of the extended network regulate each other as found in CODEX, with a tick indicating that the TF binds to a regulatory region of the corresponding gene. Table 2.3 shows whether the TFs found in NIAbank (Nanog, Oct4, Sox2 and Mycn) activate or inhibit genes on the full network. Interactions found in NIAbank are present as either an up arrow for activation or a down arrow for repression.

TF Symbol	**Extended Network genes acted on**					
	Nanog	Oct4	Sox2	Mycn	Rest	Prdm14
Nanog	✓	✓	✓	✓	✓	✓
Oct4	✓	✓	✓	✓	✓	✓
Sox2	✓	✓	✓	✓	✓	✓
Mycn		✓		✓	✓	
Rest	✓				✓	
Prdm14	✓	✓	✓	✓	✓	✓

Table 2.2: Table showing interactions present in CODEX, indicted by a tick if one of the TFs of interest binds to the regulatory region of one of the genes in the GRN. Table is read down then across. **KEY:** Interactions kept as in [62]. Interactions present in CODEX but not NIA bank. Interactions present in NIAbank but not in CODEX.

We kept the core network interactions (between Nanog, Oct4 and Sox2) the same as in the reference GRN [62] as this has already been demonstrated to reproduce mESC dynamics. We removed Rex1 because it is only considered an output to the system, and therefore not necessary to describe Nanog dynamics. Interactions found in CODEX, but not NIAbank were removed as we

TF Symbol	**Extended Network genes acted on**					
	Nanog	Oct4	Sox2	Mycn	Rest	Prdm14
Nanog	↓			↓	↑	↑
Oct4	↓			↑	↑	↓
Sox2		↓	↓	↓	↓	↓
Mycn				↓		↑

Table 2.3: Table showing interactions and directions present in NIA bank. Table is read down then across, with an up arrow indicating activation of a GRN gene by a TF and down arrow indicating inhibition. **KEY:** Interactions kept as in [62]. Interactions present in CODEX but not NIA bank. Interactions present in NIAbank but not in CODEX.

do not know whether the interaction is an activation or inhibition. Conversely, interactions found in NIAbank but not CODEX were included. In the NIAbank Oct4 was found to activate Mycn and Rest, but Sox2 was found to inhibit them both. Taking Oct4-Sox2 together as a heterodimer, with one half activating and the other inhibiting, there would be no resulting effect. Therefore no interaction of Oct4-Sox2 was included on Mycn and Rest.

The additional medium signalling interactions were also added: Chiron's activation of β-catenin, modelled as an inhibition of its degradation; the inhibition of Tcf3 by β-catenin, modelled as an activation of Tcf3's degradation; and Tcf3 inhibiting the regulation of Oct4-Sox2 on Nanog.

We used a recently published data set [168] to identify interactions of β-catenin with the core network genes and additional genes found. This data set contains microarray experiment results finding genes differentially expressed between Chiron (β-catenin activator) and XAV (β-catenin inhibitor). From this β-catenin was found to inhibit Mycn and Rest. Finally, we added protein interactions from the literature for a positive feedback loop between Nanog and Rest [140], and Prdm14 inhibition of Fgf/Erk signalling [101].

2.3.6 Summary of final network

The complete network is represented in Figure 2.18. In summary:

- The heterodimer Oct4-Sox2 represses Prdm14. It also activates itself and Nanog, the latter interaction being inhibited by Tcf3.
- Nanog activates itself, Prdm14 and Rest. It also inhibits Mycn.
- Mycn inhibits itself and activates Prdm14.
- Rest activates Nanog and Oct4-Sox2.
- Prdm14 inhibits FGF4/Erk inhibition of Nanog's auto-activation.
- FGF4/Erk is activated by Oct4-Sox2 and inhibits Nanog's positive auto-regulation.
- Chiron activates β-catenin, β-catenin inhibits Mycn, Rest and Tcf3; and Tcf3 inhibits the activation of Nanog by Oct4-Sox2.
- PD inhibits FGF4/Erk inhibition of Nanog's auto-activation.

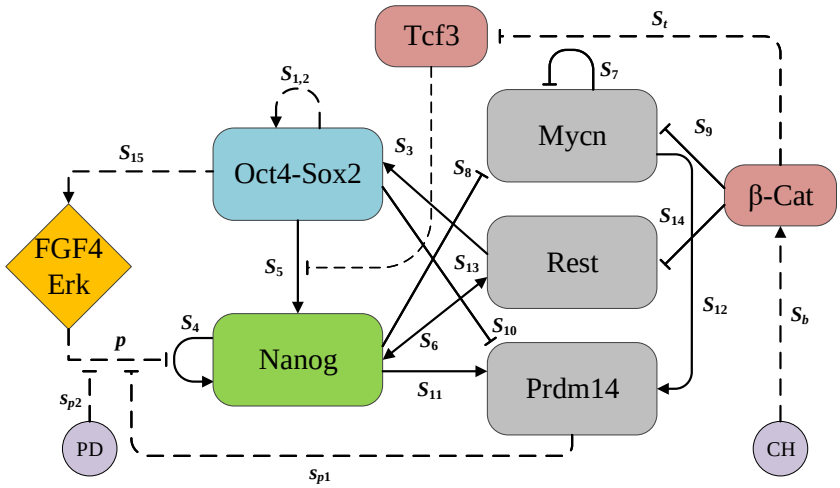


Figure 2.18: Extended GRN. Solid lines are transcriptional interactions, dashed lines are non-transcriptional interactions. Arrows are activations, flat headed arrows are inhibitions. Core network is made of Nanog, Oct4 and Sox2. Mycn, Rest and Prdm14 are the genes added which are differentially expressed between Serum+LIF and 2i+LIF. FGF4/Erk, Tcf3 and β-catenin are added as targets of the two inhibitors present in 2i+LIF: PD and Chiron (CH). Parameters correspond to equations (2.39) to (2.47).

The equations describing the GRN of Figure 2.18 are given by the following:

$$p = \frac{s_{p1}\cdot\ k_{p1}}{k_{p1}+[PR]} + \frac{s_{p2}\cdot\ k_{p2}}{k_{p2}+[PD]}, \tag{2.39}$$

$$\frac{d[OS]}{dt} = n\cdot\left(\frac{s_{1,2}\cdot\ [OS]^2}{(k+[OS])^2\cdot\ dO\cdot\ dS} + \frac{s_3\cdot\ [R]}{k+[R]} - d_{OS}\cdot\ [OS]\right) + \varphi(0,\sigma_{OS})\cdot[OS], \tag{2.40}$$

$$\frac{d[N]}{dt} = n\cdot\left(\frac{s_4\cdot\ [N]^2}{k^2+[N]^2+p\cdot\ [E]} + \frac{s_5\cdot\ [OS]}{k+[OS]+[TC]} + \frac{s_6\cdot\ [R]}{k+[R]} - d_N\cdot\ [N]\right) + \varphi(0,\sigma_N)\cdot[N], \tag{2.41}$$

$$\frac{d[M]}{dt} = n\cdot\left(\frac{s_7\cdot\ k}{k+[M]} + \frac{s_8\cdot\ k}{k+[N]} + \frac{s_9\cdot\ k_m}{k_m+[B]} - d_M\cdot\ [M]\right) + \varphi(0,\sigma_M)\cdot[M], \tag{2.42}$$

$$\frac{d[PR]}{dt} = n\cdot\left(\frac{s_{10}\cdot\ k}{k+[OS]} + \frac{s_{11}\cdot\ [N]}{k+[N]} + \frac{s_{12}\cdot\ [M]}{k+[M]} - d_{PR}\cdot\ [PR]\right) + \varphi(0,\sigma_{PR})\cdot[PR], \tag{2.43}$$

$$\frac{d[R]}{dt} = n\cdot\left(\frac{s_{13}\cdot\ [N]}{k+[N]} + \frac{s_{14}\cdot\ k_r}{k_r+[B]} - d_R\cdot\ [R]\right) + \varphi(0,\sigma_R)\cdot[R], \tag{2.44}$$

$$\frac{d[E]}{dt} = n\cdot\left(\frac{s_{15}\cdot\ [OS]}{k+[OS]} - d_E\cdot\ [E]\right), \tag{2.45}$$

$$\frac{d[B]}{dt} = n\cdot\left(\alpha_1 - d_B\cdot\left(\frac{s_b\cdot\ k_b}{k_b+[CH]}\right)\cdot\ [B]\right) + \varphi(0,\sigma_B)\cdot[B], \tag{2.46}$$

$$\frac{d[TC]}{dt} = n\cdot\left(\alpha_2 - d_{TC}\cdot\left(1+\frac{s_t\cdot\ [B]}{k_t+[B]}\right)\cdot\ [TC]\right) + \varphi(0,\sigma_T)\cdot[TC], \tag{2.47}$$

N=Nanog; OS=Oct4-Sox2; M=Mycn; PR=Prdm14; R = Rest; E=Fgf4/Erk; B=β-catenin; TC=Tcf3. Maximum transcription rates are denoted by S_i, Michaelis–Menten constants by k_i, noise amplitude by σ_i (via normally distributed noise terms φ), degradation rates by d_i, constant transcription rates by α_i and a time normalization factor of n.

Parameter values for this system can be found in Table 2.6. No noise was added to Fgf/Erk signalling (equation (2.45)) matching the reference model.

2.3.7 Parameter values and fitting

We used the principles set out in Section 2.2.6 to alter the parameters for desired system behaviour. Increasing a maximum transcription rate s_i can be compensated for by simultaneously increasing the corresponding degradation rate d_i. Therefore all degradation rates were fixed to 1, as in the reference model [62], which decreases the number of parameters that need fitting. All k_i values were fixed except where the system needed more or less sensitivity to retain a comparable order of magnitude of changing network factors. Parameters s_i were chosen such that in the

deterministic model the newly added genes have matching fold-change directions between the two media as in [83] (see below).

A time normalization factor was included to set the system to the correct time scale. This was chosen to match the half life of Nanog in the system equations to experimental measurements of its degradation. We did this by considering only Nanog's degradation with time normalization factor n, given by

$$\frac{d[N]}{dt} = -n \cdot [N]. \tag{2.48}$$

As the degradation terms d_i were all set to 1, d_N is not included here. This differential equation has a solution given by

$$[N(t)] = C \cdot \ e^{-nt}, \tag{2.49}$$

where C is a constant. The half life is the time at which $e^{-nt} = \frac{1}{2}$, which is the term

$$t_{\,half} = \frac{ln2}{n}. \tag{2.50}$$

We chose a value of $n = 0.005$ resulting in the half life of Nanog in the system being 138.63. If the units of the system are in minutes this is equivalent to 2.31 hours, matching the measured half life of Nanog in [3] (stated as 2.3 hours). This is applied to all the genes in the system meaning they all have half-lives set the same as Nanog. As Nanog is the main focus of the system and having only one time normalization factor keeps the number of parameters to a small number, this simplification is justified.

We found noise parameters σ_i by first finding a noise value in serum+LIF for Nanog which gave a distribution of approximately half NH and half NL (based on Figure 2.16) and setting the rest of the noise parameters to around half of this value to be consistent with measurements of Nanog having greater fluctuations [4]. Nanog's maximum transcription rate s_4 was also simultaneously altered at this stage to output the desired distribution (based on Figure 2.16). Final parameter values of the equations in (2.39) to (2.47) can be seen in Table 2.6. Values for PD and Chiron (CH) are changed depending on which culture conditions are being simulated, both set at zero for Serum+LIF and increased for 2i+LIF. Two different parameter value sets were used for simulating 2i+LIF (see Section 2.4.2 for more details); PD and Chiron both set to 2 arbitrary units (AU) with Nanog transcriptional noise reduced by 20% and PD and Chiron both set to 20 AU.

The deterministic system for PD, CH=0 AU and PD, CH=2 AU is bistable (see Figures 2.19a and 2.24a), with two steady-states corresponding to NH and NL. As Nanog is the master pluripotency gene, all the other genes in the deterministic system have concentration levels tied to Nanog, i.e. each gene has a different deterministic steady-state for NH and NL respectively. For PD, CH=20 the deterministic system is monostable, with only one steady-state set of concentrations corresponding to NH. The steady-state concentration values of the newly added genes for NH and NL cell states and for each of the different PD and CH value scenarios (only NH reported for PD, CH =20) are shown in Table 2.4.

This is only the steady-state value for a single cell. To estimate concentration levels of the new genes in the different media, we considered the proportion of cells reported in the literature expressing NH or NL. In Serum+LIF, around 50% of mESCs have a NL concentration and 50% NH [3, 4]. The steady-state concentrations of the newly added genes in Serum+LIF was therefore calculated as

$$C_{serum} = 0.5 \cdot SS_{NH} + 0.5 \cdot SS_{NL}, \tag{2.51}$$

where C_{serum} is the calculated concentration in Serum+LIF, SS_{NH} is the steady-state value corresponding to NH and SS_{NL} is the steady-state value corresponding to NL. In 2i+LIF approximately 90% of cells are in a NH state and 10% in the NL [3, 4]. Concentrations for 2i+LIF were calculated as

$$C_{2i} = 0.9 \cdot SS_{NH} + 0.1 \cdot SS_{NL}, \tag{2.52}$$

where C_{2i} is the calculated concentration in 2i+LIF. For PD and CH=20, with no NL state, the values from the PD and CH=2 for NL concentrations were used. Fold change between the two media is calculated as

$$log_2\left(\frac{C_{2i}}{C_{serum}}\right). \tag{2.53}$$

We simulated the deterministic system, recorded the steady-state values and calculated the fold change. This was done iteratively until the fold change directions of the deterministic system best matched the data of [83]. Table 2.5 shows the final concentrations calculated from the deterministic model in each of the media, and the fold changes between 2i+LIF and Serum+LIF. The table also shows the reported experimental fold changes; although the actual values differ, the signs (positive or negative) are the same. Qualitatively, the new genes are reported higher or lower in Serum+LIF versus 2i+LIF in the deterministic model the same as in the data set.

Gene	NL steady-state value	NH steady-state value
	PD, Chiron=0	
Mycn	100.5	39.3
Prdm14	15.8	33.4
Rest	14.3	14.4
	PD, Chiron=2	
Mycn	92.6	34.2
Prdm14	16.6	33.2
Rest	9.1	9.2
	PD, Chiron=20	
Mycn	N/A	27.5
Prdm14	N/A	32.9
Rest	N/A	2.2

Table 2.4: Table showing steady-state concentrations for newly added genes (deterministic system only). Each gene has a seperate steady-state corresponding to NH and NL.

Gene	Serum+LIF concentration	2i+LIF concentration (PD, CH=2)	2i+LIF concentration (PD, CH=20)
Mycn	69.9	40.0	34.0
Prdm14	24.6	31.6	31.3
Rest	14.4	9.2	2.9
	Fold change (PD, CH=2)	**Fold change (PD, CH=20)**	**Fold change (Data [83])**
Mycn	-0.81	-1.04	-1.71
Prdm14	0.36	0.35	1.49
Rest	-0.65	-2.31	-1.08

Table 2.5: Table showing fold change of deterministic system for newly added genes, with comparison to fold change data values taken from [83].

2.3.8 Simulation and bifurcation analysis

Numerical solutions of the SDEs were performed using the Euler-Maruyama method implemented using the MATLAB platform (Mathworks). A total time of 5760 equates to four days of simulation. Step-size $\Delta\, t$ of each update step was 0.01, equivalent to 0.6 seconds. Bifurcation analysis (i.e. numerical continuation) was performed on the deterministic system (no noise) using the software xppaut.

s_{p1}	30*	s_{14}	20*	d_M	1#
s_{p2}	30*	s_{15}	2*	d_{PR}	1#
k_{p1}	5*	s_b	50*	d_R	1#
k_{p2}	15*	s_t	20*	d_E	1#
$s_{1,2}$	75*	α_1	1*	d_B	1#
s_3	1*	α_1	40*	d_{TC}	1#
s_4	37*	k	10	σ_{OS}	0.02
s_5	2*			σ_N	0.05 (0.04 for PD=CH=2)
s_6	0.01*	k_r	0.05	σ_M	0.02
s_7	10*	k_m	0.05	σ_{PR}	0.02
s_8	100*	k_t	10	σ_R	0.02
s_9	20*	k_b	1	σ_B	0.02
s_{10}	15*	d_O	1#	σ_T	0.02
s_{11}	30*	d_S	1#	s_{nb}	10*
s_{12}	10*	d_{OS}	1#	n	0.005
s_{13}	0.1*	d_N	1#		

(* molecules/minute; # 1/minute)

Table 2.6: Table showing parameter values for full network described in equations (2.39) to (2.47).

2.4 New model behaviour/results

We fitted the model to the distributions of Figure 2.2b, for Serum+LIF and 2i+LIF scenarios respectively. We altered the parameters s_4 and σ_N such that the proportions of NH and NL matched between the model and the data. We used a simple threshold to define NH and NL, with any cell expressing a Nanog level lower than the threshold characterized as NL, and as NH for greater than the threshold. The threshold value we used was 5, (equivalent to $10^{0.6990}$ when referring to simulated distribution results).

2.4.1 Serum+LIF results

For simulating Serum+LIF culture conditions we set the values of PD and Chiron (CH) to 0 AU. We explored the model with continuation analysis on the deterministic system (no noise) to see the stability of the system. Figure 2.19a shows a continuation of the steady-state of Nanog against its auto-regulatory rate s_4. The blue part of the curve denotes stable steady-state, whilst the red part is unstable. The points at which a blue curve meet a red curve are Saddle-node bifurcations. The chosen value of s_4 is represented by a dashed black line which intersects a lower stable steady-state (NL steady-state) and a higher stable steady-state (NH steady-state) separated by an unstable steady-state. The deterministic system therefore exhibits bistability. Adding the stochastic elements to the deterministic system (resulting in the full SDE shown in equations (2.39) to (2.47)) allows it to switch between the NL and NH stable steady-state.

We simulated the SDE model 10,000 times to represent individual instances of mESCs and recorded the concentration of the cells after steady-state was reached. This distribution is shown in Figure 2.19b which shows a bistable distribution of Nanog, matching experimental results shown in Figure 2.2b of approximately 50% of cells in NH and NL states respectively. Simulated distribution is shown as a histogram with fitted kernal density (red line), normalized as a probability density function. We also simulated the SDEs with Serum+LIF parameters 50 times, each with 1000 individual cells. NH was $49.6 \pm 1.5\%$ of the overall populations as compared to the proportion of NH observed experimentally of $56.2 \pm 8\%$ [3, 4].

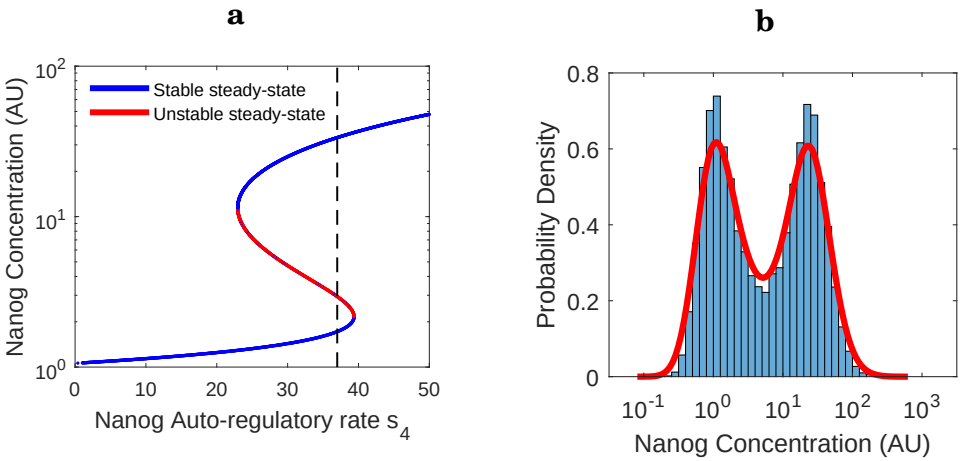


Figure 2.19: GRN network results for Serum+LIF parameters (PD=CH=0). **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects a low and a high stable steady-state, indicating bistability. **b** Distribution of Nanog's steady-state concentration in 10,000 simulated cells, showing approximately 50% of the cells are in the NL state and 50% in the NH state.

Flow cytometry sorting experiment results from [3] are shown in Figure 2.20b. The authors measured the proportion of mESCs expressing NH over 4 days, with three different starting populations: a pure NH population, a pure NL population and a mixed population (unsorted mESCs). We simulated the SDE model for 4 days, once with starting conditions of all NH and once with all NL (1000 simulations for each starting state). The simulations shown in Figure 2.20a correctly match the experimental observations of both sub populations, NH and NL, restoring bistable distributions within 4 days of culture.

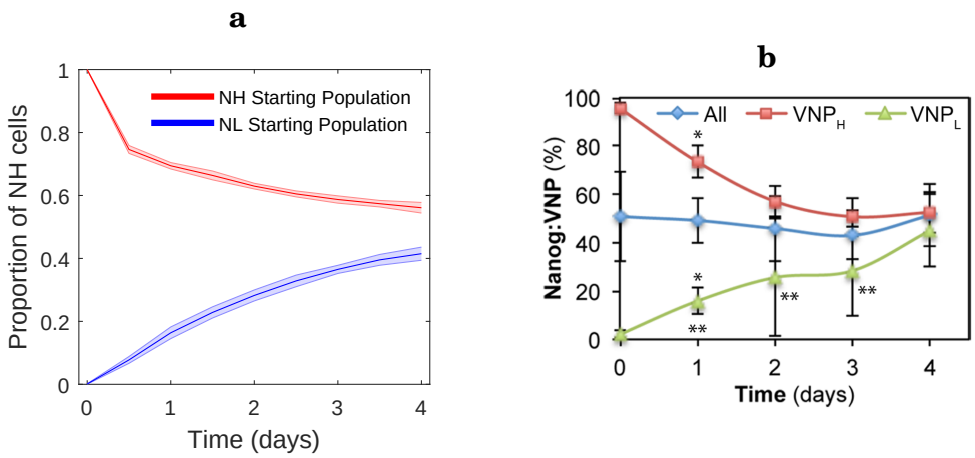


Figure 2.20: Sorting of cells in Serum+LIF. **a** Simulation of cells sorting (PD=CH=0). Two simulations were carried out, one with pure NH cells and the other with only NL. Over time the distirubtions of each of these starting populations tended towards the steady-state of approximately 50% NH and 50% NL. **b** Sorting of Nanog from experimental results, adapted from [3]. VNP_H and VNP_L are starting populations of pure NH and NL cells respectively. This qualitavely matches our simulation, with both populations tending towards the same steady-state distribution.

The concentration of Nanog and Oct4-Sox2 over 15 hours is shown for four representative single cell simulations in Figure 2.21. Nanog is observed to stochastically fluctuate, whilst Oct4-Sox2 remains steady. Black dashed line represents threshold value used to separate NH from NL. This matches experimental single-cell measurements of Nanog in [4], shown in Figure 2.22 (experimental time courses are for Nanog only), which show fluctuations between NH and NL within circa 15 hours.

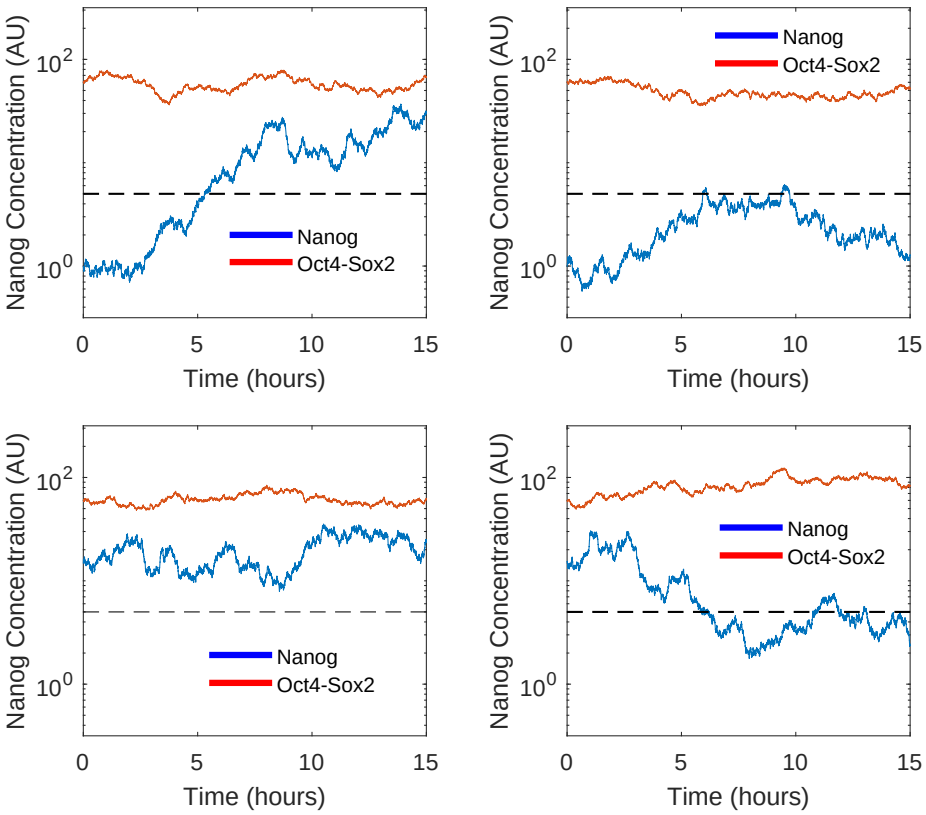


Figure 2.21: Four representative time courses of Nanog and Oct4-Sox2 concentration in four cells over 15 hours for Serum+LIF parameter values (PD, CH=0). Dashed black line is threshold we used to separate NH and NL cells.

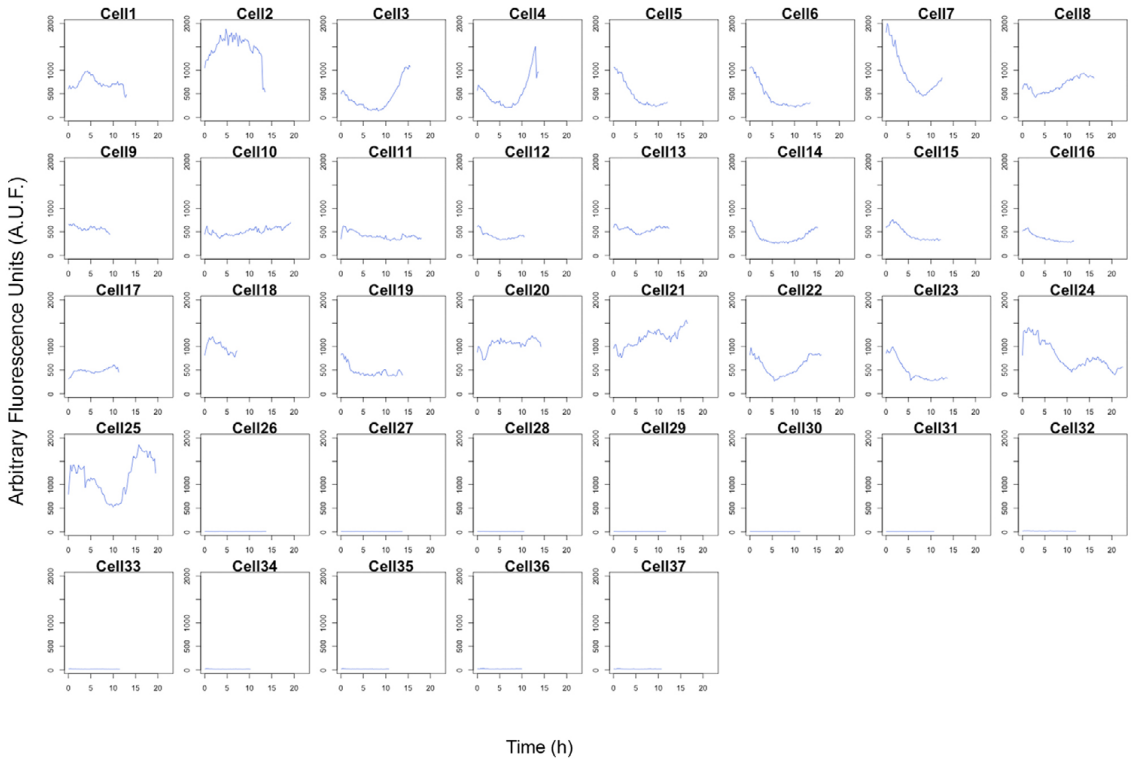


Figure 2.22: Experimental data (time lapse) of Nanog measured in individual cells cultured in Serum+LIF, adapted from [4]. Individual cells swap between NH and NL states within a cell cycle (15 hours). Blank data was from dead cells or errors in the recording.

2.4.2 2i+LIF results

For simulating 2i+LIF culture conditions, we considered two different parameter cases. For the first case considered, the values of PD and Chiron were simultaneously increased until the distributions matched the desired proportions of approximately 10% of cells expressing NL and 90% NH. The value we set both PD and Chiron to achieve this was 20 AU. Figure 2.23a shows the continuation of Nanog's steady-state against its auto-regulatory rate s_4, with the dashed black line denoting its value. Here the deterministic system is fully monostable, shown by the black dashed line intersecting only the NH state. We ran the SDEs with these 2i+LIF parameter for 10,000 cells, resulting in the distribution shown in Figure 2.23b. A small population of NL cells is still present, even though the deterministic system no longer has a NL steady-state. Even with the deterministic system being fully monostable, the stochasticity of the system is enough to drive some of the cells to the NL state. Another element of the system attracting cells towards the NL state is the ghost bottleneck, described in [141]. This is where systems close to Saddle-node bifurcations have longer transients as the system passes underneath. We simulated the SDEs with these 2i+LIF parameters (PD, CH=20) 50 times, each with 1000 individual cells. NH was $87.27 \pm 0.99\%$ of the overall populations as compared to the proportion of NH observed experimentally of $91.1 \pm 3.1\%$ [3, 4].

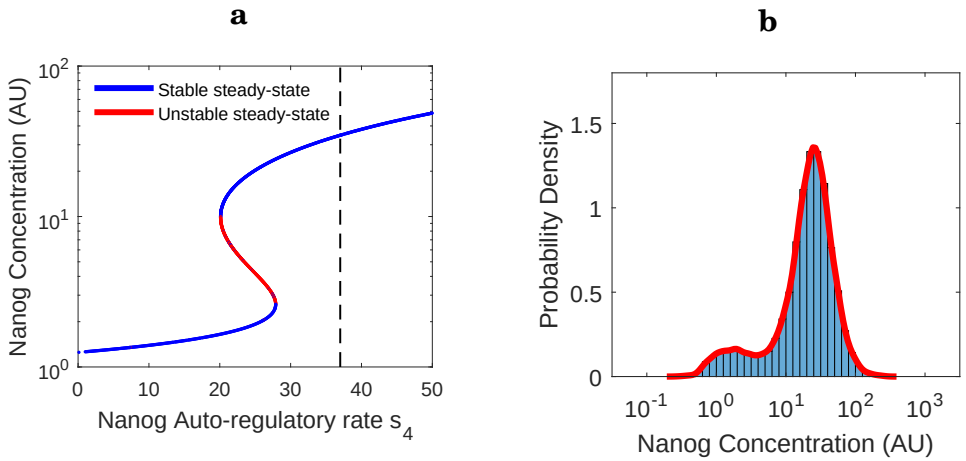


Figure 2.23: GRN network results for 2i+LIF parameters (PD=CH=20).
a Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects only the high steady-state, indicating a shift in the deterministic system towards monostability.
b Distribution of Nanog's steady-state concentration in 10,000 simulated cells, showing approximately 90% of the cells in the NH state and 10% in the NL state.

The second 2i+LIF parameters case we explored is increasing PD and Chiron, but also decreasing Nanog's stochastic noise. This is more consistent with experimental results which measured the number of mRNA molecules per cell showing that the fluctuations of Nanog are

decreased in 2i+LIF [4]. The specific case we considered here is PD and Chiron equal to 2 and Nanog's noise term σ_N reduced by 20% from 0.05 to 0.04. The continuation of Nanog's steady-state against s_4 for this 2i+LIF scenario can be seen in Figure 2.24a. The system is still bistable, though the value of s_4 (black dashed line) is now very close to the Saddle-node bifurcation. In combination with reducing the noise, the SDE model can match the observed distribution of Nanog in 2i+LIF of approximately 90% expressing NH and 10% NL (Figure 2.24b). We again simulated the SDEs with these new 2i+LIF parameters (PD, CH=2; noise reduced by 20%) 50 times, each with 1000 individual cells. NH cells were $88.46 \pm 1\%$ of the overall populations as compared to the same proportion of NH observed experimentally of $91.1 \pm 3.1\%$ [3, 4]. Therefore both of these 2i+LIF scenarios give qualitatively the same distributions of Nanog.

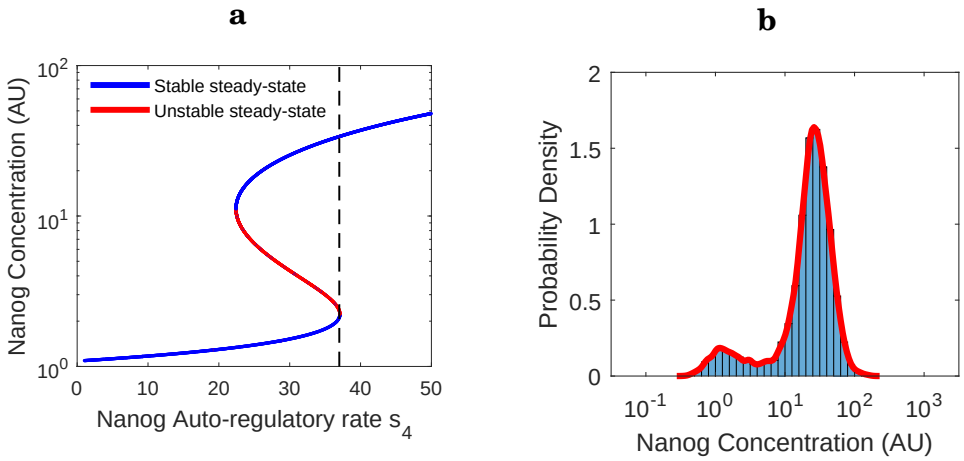


Figure 2.24: GRN network results for 2i+LIF parameters (PD=CH=2, noise reduced 20% [4]). **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects both the high and low steady-state, though is close to the Saddle-node bifurcation of the low state. The system is still bistable, though has shifted towards monostability. **b** Distribution of Nanog's steady-state concentration in 10,000 simulated cells, showing approximately 90% of the cells in the NH state and 10% in the NL state.

We ran flow cytometry sorting simulations for both 2i+LIF parameter sets, with the SDEs run for four simulated days with starting populations of pure NH and NL cells (1000 realizations for each). Figure 2.25a is for PD and Chiron equal to 20, and Figure 2.25b is for PD and Chiron equalling 2, with Nanog noise reduced 20%. For both 2i+LIF cases the cells tend towards the steady-state distribution of approximately 90% NH, 10% NL. However, simulated cells for the first case tend towards this steady-state much faster. Sorting experiments in 2i+LIF have not been performed, and could inform which case is a better representative of mESCs in 2i+LIF.

Four representative cell time courses of Nanog and Oct4-Sox2 concentration for PD=Chiron=20 scenario are shown in Figure 2.26. Oct4-Sox2 concentration remains stable over the 15 hours,

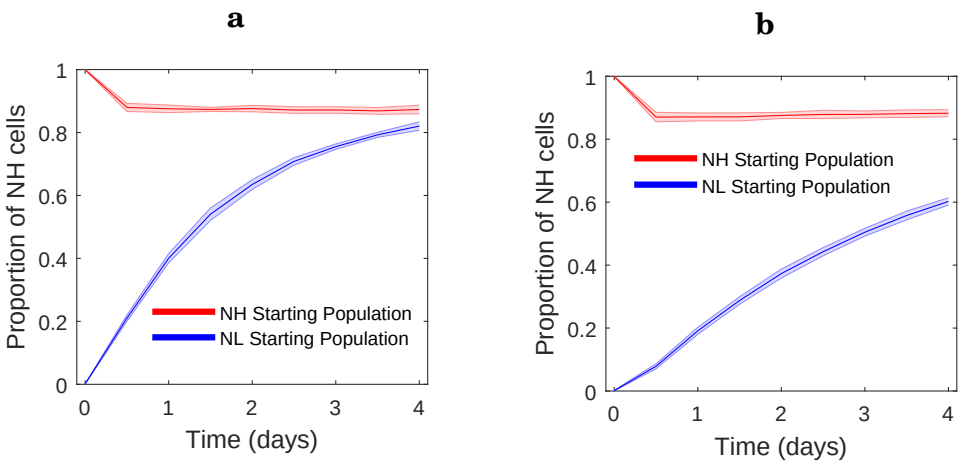


Figure 2.25: Sorting of cells in 2i+LIF. **a** Simulation of cells sorting (PD=CH=20). Two simulations were carried out, one with pure NH cells and the other with only NL. Over time the distributions of each of these starting populations tended towards the steady-state of approximately 90% NH and 10% NL. **b** Simulation of cells sorting (PD=CH=2 and noise reduced 20%). Over time the distributions of both starting populations tends towards the stable steady-state of approximately 90% NH and 10% NL. However they approach this steady-state slower.

whilst Nanog concentration fluctuates. Black dashed lines represent threshold value used to separate NH from NL.

Four representative cell time courses of Nanog and Oct4-Sox2 concentration for PD=Chiron=2 and noise reduced 20% scenario are shown in Figure 2.27. Similar to the PD, Chiron=20 case, cells express stable concentrations of Oct4-Sox2, with fluctuating levels of Nanog. Black dashed lines represent threshold value used to separate NH from NL. Qualitatively, simulation time courses in Figures 2.26 and 2.27 match experimental single-cell measurements in Figure 2.28 (experimental time courses are for Nanog only).

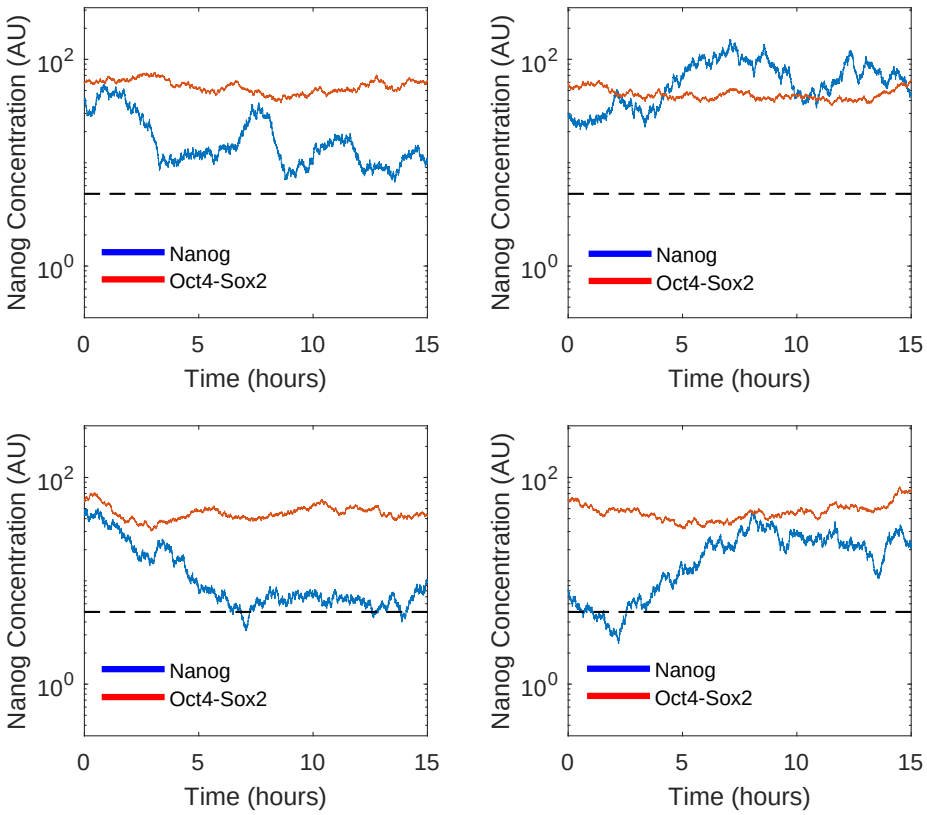


Figure 2.26: Four representative time courses of Nanog and Oct4-Sox2 concentration in four cells over 15 hours for 2i+LIF parameter values (PD, CH=20). Dashed black line is threshold we used to separate NH and NL cells.

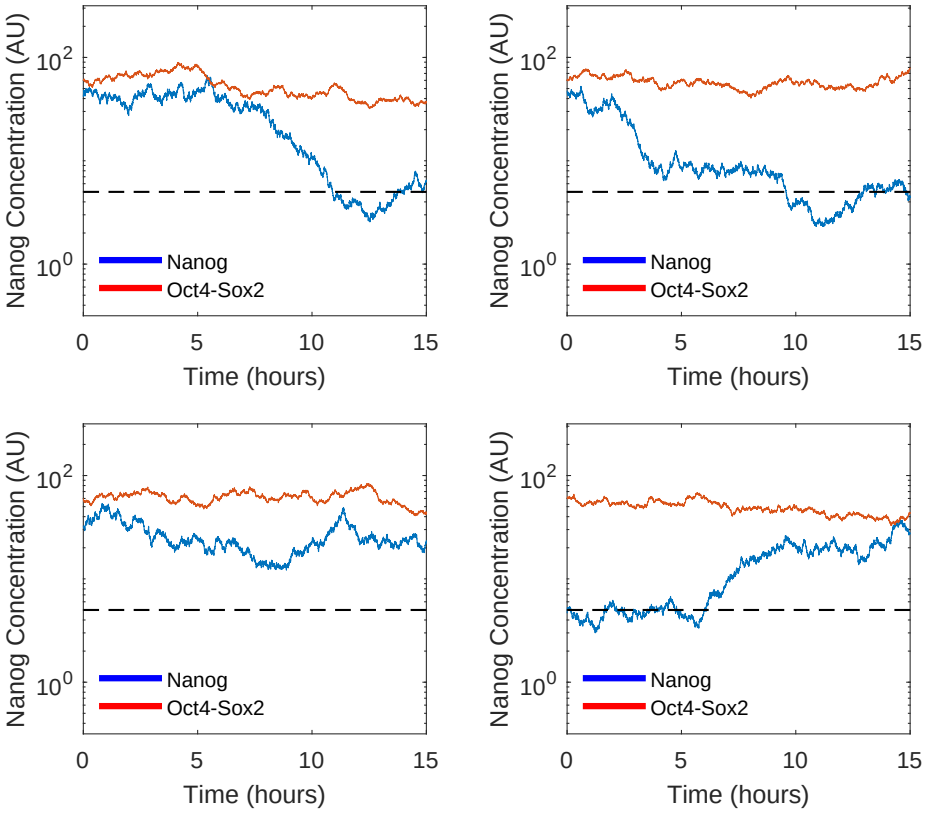


Figure 2.27: Four representative time courses of Nanog and Oct4-Sox2 concentration in four cells over 15 hours for 2i+LIF parameter values (PD, CH=2; noise reduced 20%). Dashed black line is threshold we used to separate NH and NL cells.

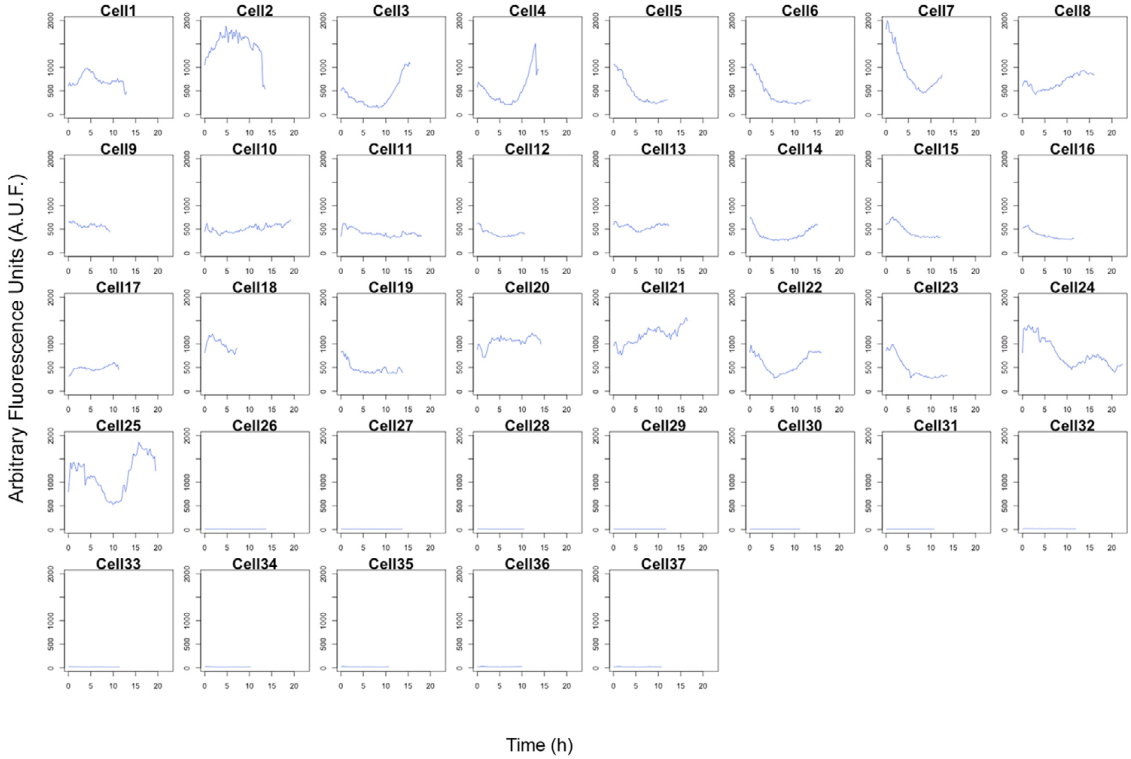


Figure 2.28: Time lapse of Nanog measured in individual cells cultured in 2i+LIF, adapted from [4]. Blank data was from dead cells or errors in the recording.

Of note, in 2i+LIF simulation results shown hereafter, the reduced noise parameters were used. This is to more realistically represent experimental data. Furthermore, distribution plots are now normalized as % of Max, and only kernal fitting functions are shown.

2.4.3 Gene/molecular pathway deletions

To understand how the newly added genes effect the system stability, we performed 2-parameter continuations on the variables PD and Chiron. This was for the whole network, and under single gene deletions. To simulate a gene's deletion we set all terms contributing towards that gene in the differential equation system to 0, and the initial concentration of that gene was also set to 0, ensuring that the gene's concentration remains at nothing for all time. Figure 2.29 shows the 2-parameter continuations of the whole network, and upon single gene deletions. We performed this in xppaut, continuing a Saddle-node bifurcation as found from single parameter continuations. The area above each respective curve denotes the region of PD and Chiron concentrations for which the deterministic system is monostable, NH only.

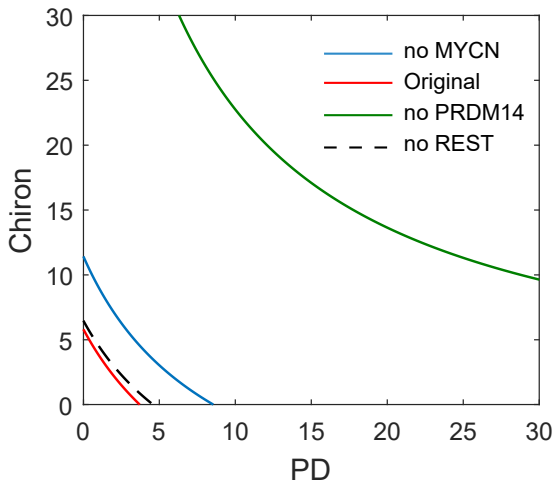


Figure 2.29: 2-parameter continuations of the Saddle-node from Nanog's continuation against PD and Chiron; with the full network, and with gene deletions. Each curve is a Saddle-node bifurcation, above which the system is monostable.

For the whole network (Figure 2.29, red line) both inhibitors enable the shift of the system towards a single NH steady-state, PD via altering Nanog's positive feedback regulation and Chiron via strengthening Oct4-Sox2 activation of Nanog (indirectly the same mechanism as in Figure 2.11). Deletion of Rest (Figure 2.29, black dashed line) does not significantly affect the systems stability, whilst for Mycn deletion a monostable NH state can only be obtained if PD and Chiron are further increased (Figure 2.29, blue line). This is due to the indirect regulation of Mycn on Nanog through Prdm14. The effect is much stronger upon Prdm14 deletion, requiring a

significant increase of PD and Chiron (Figure 2.29, green line).

The results from the 2-parameter continuations prompted further exploration of the effects of deleting the newly added genes, especially for Prdm14. We therefore simulated the SDE model, deleting each of the genes, to see what effect this would have on the steady-state distribution of Nanog. Upon Rest and Mycn deletion no change in the steady-state distribution of Nanog was observed (they appear the same as in Figures 2.19b and 2.24b for Serum+LIF and 2i+LIF parameter values respectively). This shows how important noise is in the system, as the noise is able to maintain the distributions of Nanog, even with changing of the deterministic steady-states. However, upon Prdm14 deletion the steady-state distribution of Nanog was significantly changed. Figure 2.30 shows steady-state distributions of Nanog simulated in 1000 cells with Prdm14 deleted. Figure 2.30a is for Serum+LIF and shows a greater proportion of mESCs expressing NL, implying a large reduction in pluripotency (compare to Figure 2.19b). For 2i+LIF, shown in Figure 2.30b, Nanog heterogeneity persists (compare to Figure 2.24b). This is consistent with experimental results showing that upon Prdm14 deletion mESCs cultured in 2i+LIF display heterogeneous Nanog expression and defective differentiation potential [164].

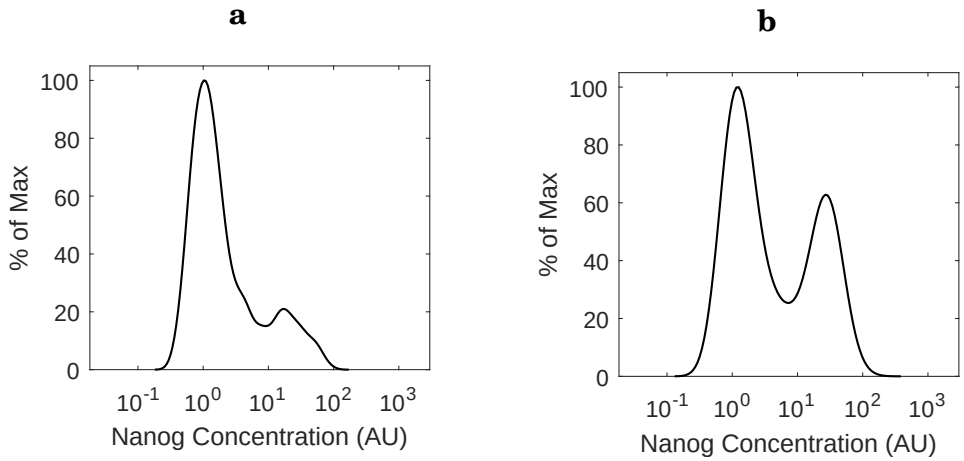


Figure 2.30: Distriubtions of Nanog upon Prdm14 deletion for **a** Serum+LIF parameters and **b** 2i+LIF parameters. Both parameter sets show a significant increase in the proportion of NL mESCs, indicating a reduction of pluripotency.

To explore what effect β-catenin and Tcf3 have on the stability of the system, we performed continuations of the steady-state of Nanog against its auto-regulatory rate upon β-catenin deletion and Tcf3 deletion. When β-catenin is deleted, the system is pushed further into the bistable regime for both Serum+LIF (Figure 2.31a) and 2i+LIF (Figure 2.31b) as shown by the dashed black lines; compare to Figures 2.19a and 2.24a respectively.

More pronounced effects are seen for Tcf3 deletion. For Serum+LIF parameters (Figure 2.32a) the system has shifted entirely into a NH mono stable regime, with a further shift for 2i+LIF parameters (Figure 2.32b) as compared to Figures 2.19a and 2.24a respectively. This is consistent

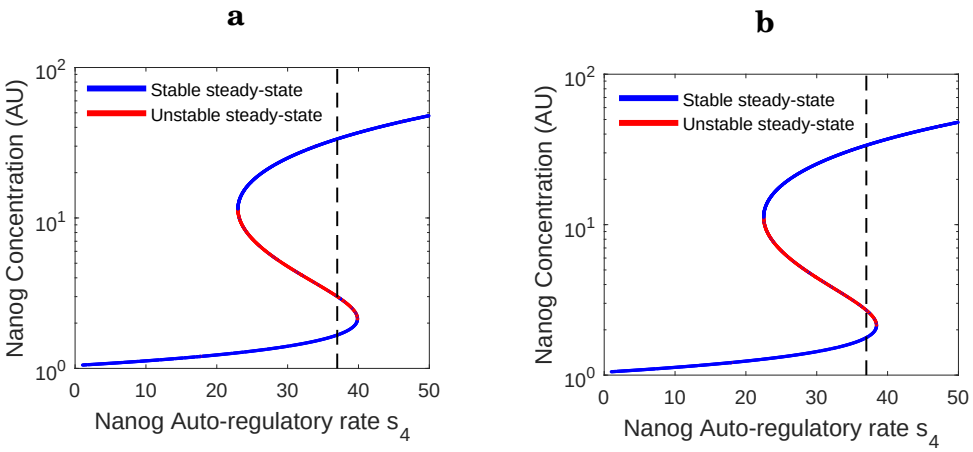


Figure 2.31: Continuations of Nanog steady-state upon β-catenin deletion, in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. **a** Serum+LIF parameters. The current transcription rate s_4 (black dashed line) intersects both the high and low steady-state, indicating bistability. **b** 2i+LIF parameters. The current transcription rate s_4 (black dashed line) intersects both the high and low steady-state, though has moved closer to the Saddle-node bifurcation. The system is slightly further into the bistable regime upon β-catenin deletion.

with experimental results showing that when Tcf3 is inhibited, mESC differentiation is delayed [165], i.e. the cells are more pluripotent.

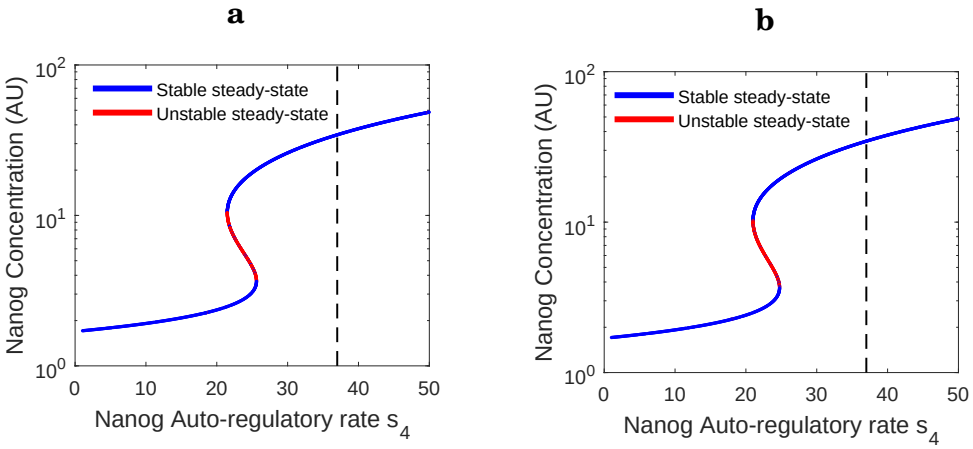


Figure 2.32: Continuations of Nanog steady-state upon Tcf3 deletion. **a** Serum+LIF parameters. The current transcription rate s_4 (black dashed line) intersects only the high steady-state. **b** 2i+LIF parameters. The current transcription rate s_4 (black dashed line) intersects only the high steady-state, though is further into the monostable regime. Both parameter sets show a significant shift of the system into the high monostable regime upon Tcf3 deletion.

2.4.4 Alternative network structures/modifications to the GRN

Nanog has been reported to indirectly activate β-catenin [87], inhibiting Dkk1 which inhibits β-catenin. This interaction was reported to be important for successfully reprogramming fibroblast cells (a type of cell that creates the structure for animal tissues) into pluripotent stem cells. Therefore, we explored this interaction by adding it to the model. For simplicity this was modelled as a direct activation by Nanog, ignoring the double inhibition through Dkk1. The part of the system equations altered with this extra activation (equation (2.46)) is now given by

$$\frac{d[B]}{dt} = n \cdot \left(\alpha_1 + \frac{s_{nb} \cdot [N]}{k + [N]} - d_B \cdot \left(\frac{s_b \cdot k_b}{k_b + [CH]} \right) \cdot [B] \right) + \varphi(0, \sigma_B) \cdot [B], \tag{2.54}$$

showing the addition of an s_{nb} Hill term. Parameter values are in Table 2.6.

Figure 2.33a shows the GRN interaction of the added activation, and Figure 2.33b shows 2-parameter continuations of the new system, as well as without the added interaction for comparison. This shows that adding this interaction reduces the amount of PD and Chiron (in the deterministic model) needed to push the system towards a monostable regime.

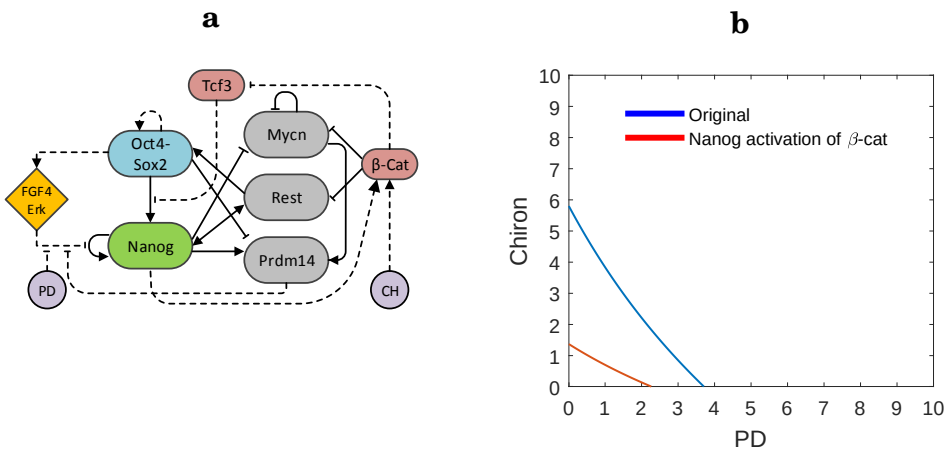


Figure 2.33: Added activation of β-catenin by Nanog. **a** Extended GRN of Figure 2.18 with added interaction of Nanog on β-catenin. **b** 2-parameter continuations of the Saddle-node from Nanog's continuation against PD and Chiron; with original network, and with added activation. Each respective curve is a Saddle-node bifurcation, above which the system is monostable.

We ran the SDE model for the original system and with the added activation shown in Figure 2.33a to compare steady-state distributions of both Nanog and β-catenin. Figure 2.34a shows the steady-state distribution of Nanog, without Nanog activating β-catenin and for Serum+LIF

parameter values. Figure 2.34b shows the steady-state distribution of β-catenin under the same conditions. The steady-state of Nanog is the same as in 2.19b demonstrating Nanog's heterogeneity with a bimodel distribution. With no activation by Nanog, β-catenin's steady-state distribution is monostable.

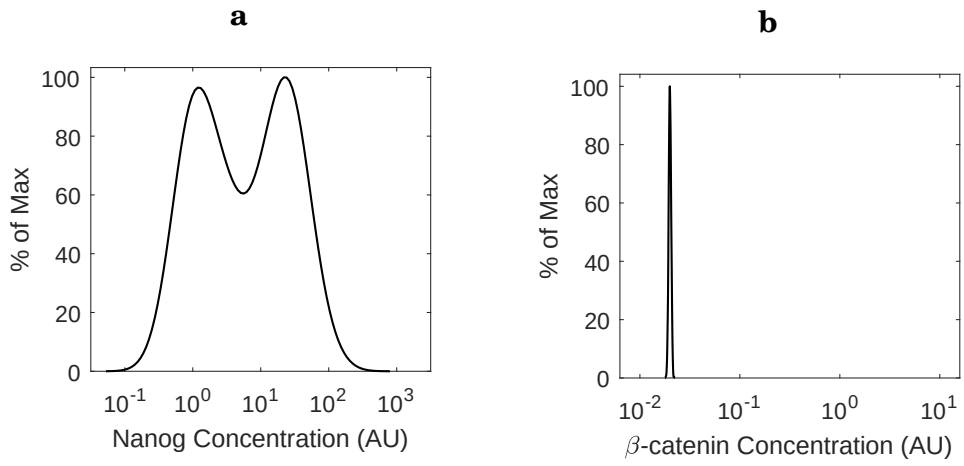


Figure 2.34: Distributions of Nanog (**a**) and β-catenin (**b**) with no activation of β-catenin by Nanog and using Serum+LIF parameters. Approximately 50% of the mESCs are in a NH state, and 50% NL. The distribution of β-catenin is monostable.

When Nanog activating β-catenin is added the steady-state of Nanog is unchanged, as shown in Figure 2.35a (compare to Figure 2.34a). This again demonstrates how the system can be more sensitive to noise than to stability of the deterministic system, as the 2-parameter continuation shows the deterministic system is affected, but this does not impact on the SDE model when noise is added. With the added activation the steady-state distribution of β-catenin is now bimodel, as shown in Figure 2.35b. The difference between the high β-catenin and low β-catenin state is smaller than for Nanog, correctly matching results as reported in [87].

Figure 2.36a shows the steady-state distribution of Nanog, without Nanog activating β-catenin and for 2i+LIF parameter values. Figure 2.36b shows the steady-state distribution of β-catenin under the same conditions. The steady-state of Nanog is the same as in 2.24b with approximately 90% of cells in NH and 10% in NL. With no activation by Nanog, β-catenin has a monostable distribution and the concentration value has been increased as compared to in Serum+LIF (compare to Figure 2.34b) due to β-catenin being directly activated by Chiron.

With 2i+LIF parameters, and now including Nanog activation on β-catenin, Nanog's steady-state distribution remains unaffected as shown in Figure 2.37a (compare to Figure 2.36a). β-catenin now has a bimodel distribution, with more cells in a high β-catenin state as compared to in Serum+LIF (Figure 2.35b).

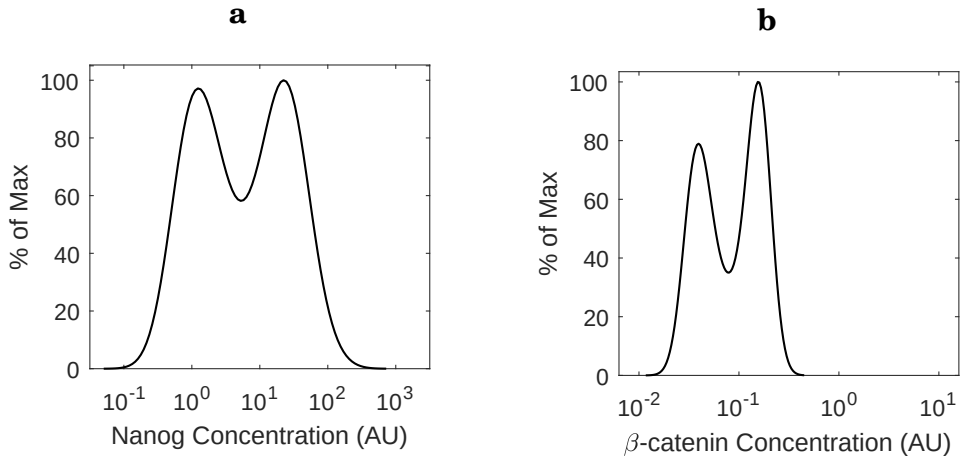


Figure 2.35: Distributions of Nanog (**a**) and β-catenin (**b**) with activation of β-catenin by Nanog and using Serum+LIF parameters. Approximately 50% of the mESCs are in a NH state, and 50% NL. The distribution of β-catenin is now bistable with just over half of the mESCs in a β-catenin high state.

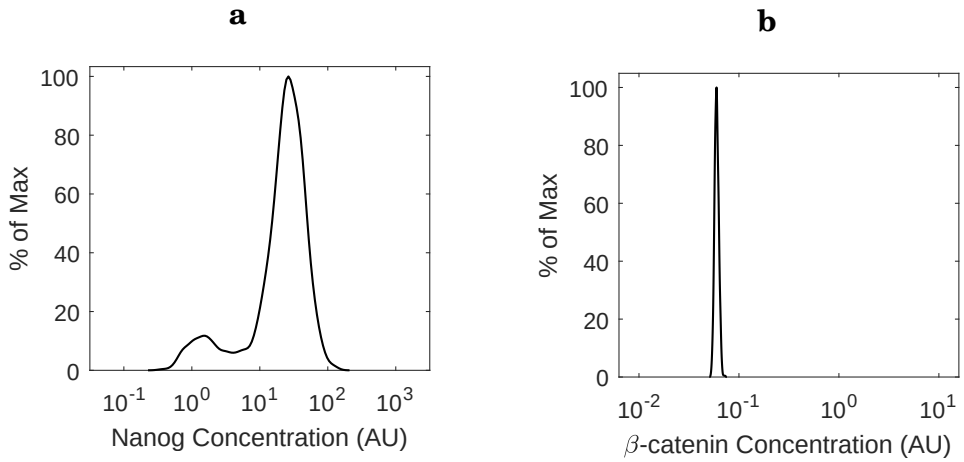


Figure 2.36: Distributions of Nanog (**a**) and β-catenin (**b**) with no activation of β-catenin by Nanog and using 2i+LIF parameters. Approximately 90% of the mESCs are in a NH state, and 10% NL. The distribution of β-catenin is monostable, and at a higher level than for Serum+LIF.

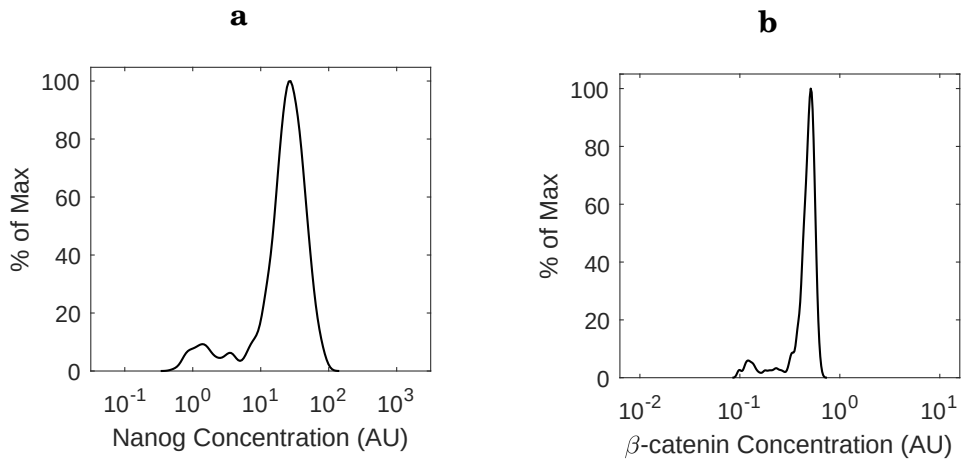


Figure 2.37: Distributions of Nanog (**a**) and β-catenin (**b**) with activation of β-catenin by Nanog and using 2i+LIF parameters. Approximately 90% of the mESCs are in a NH state, and 10% NL. The distribution of β-catenin is now bistable with approximately 90% of the cells in a β-catenin high state.

The interaction between β-catenin and Mycn has conflicting results in different experimental studies. The data set we used predicted β-catenin to inhibit Mycn. However a different study reported a positive feedback loop between Mycn and β-catenin [40]. Therefore we considered both the case of no regulation of β-catenin on Mycn, and the case of β-catenin activating Mycn. The part of the GRN equations altered to delete β-catenin regulating Mycn (equation (2.42)) is now given by

$$\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ \frac{d[M]}{dt} = n \cdot \left(\frac{s_7 \cdot k}{k+[M]} + \frac{s_8 \cdot k}{k+[N]} - d_M \cdot [M] \right) + \varphi(0,\sigma_M) \cdot [M], & & & \\ \vdots & \vdots & \vdots & \vdots \end{array} \tag{2.55}$$

where the s_9 Hill term has been removed. Parameter values are in Table 2.6

Figure 2.38a shows a schematic for the deletion of the regulation of β-catenin on Mycn. Figure 2.38b shows a 2-parameter continuation on PD and Chiron, where the area above the curve denotes the parameter region for which the system is monostable (NH only). Comparing to Figure 2.29 (Original, red curve) and Figure 2.33b (Original, blue curve) no obvious difference is observed.

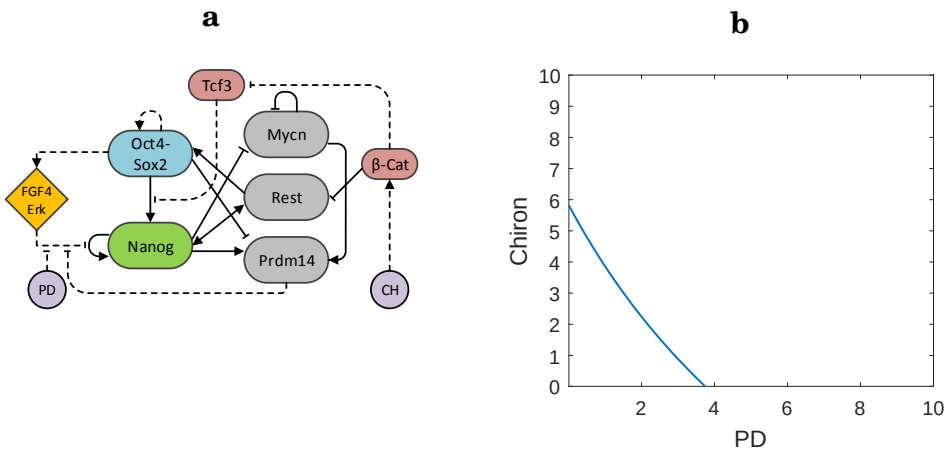


Figure 2.38: β-catenin interaction on Mycn removed. **a** Extended GRN of Figure 2.18 with interaction of β-catenin on Mycn removed. **b** 2-parameter continuation of the Saddle-node from Nanog's continuation against PD and Chiron, above the curve the system is monostable.

Continuing the steady-state of Nanog against its auto-regulatory rate s_4 when deleting this interaction resulted in no change to the stability of the deterministic system, for both medium parameters, (compare Figure 2.39a to 2.19a for Serum+LIF and 2.40a to 2.24a for 2i+LIF). The same lack of change is observed when simulating the full SDE, resulting in the same distributions

of approximately half NH and NL in Serum+LIF (compare Figure 2.39b to 2.19b for Serum+LIF) and approximately 90% NH, 10% NL in 2i+LIF (compare Figure 2.40b to 2.24b for 2i+LIF).

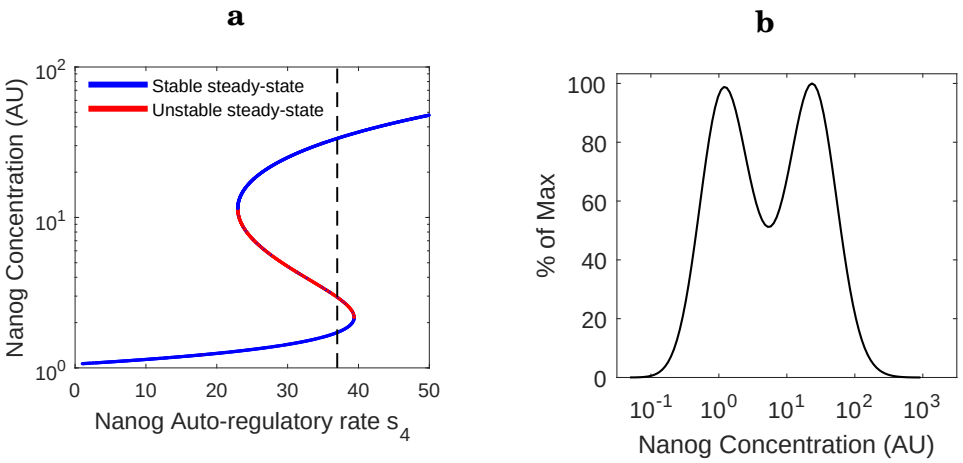


Figure 2.39: β-catenin interaction on Mycn removed results for Serum+LIF parameters. **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dahsed line) intersects both the NH and NL steady-state, indicating bistability. **b** Distribution of Nanog's steady-state concentration, showing approximately 50% of the cells in the NH state and 50% in the NL state.

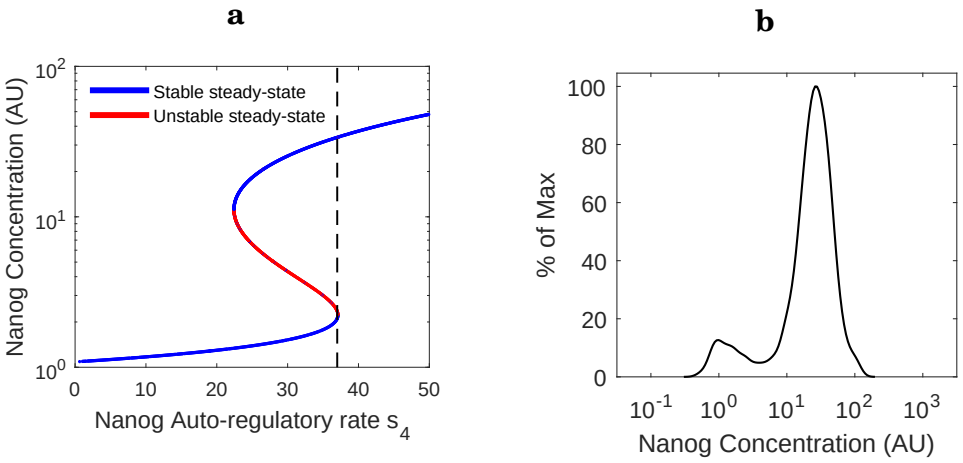


Figure 2.40: β-catenin interaction on Mycn removed results for 2i+LIF parameters. **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, though has shifted nearer to the Saddle-node bifurcation, and to the NH monostable regime. **b** Distribution of Nanog's steady-state concentration, showing approximately 90% of the cells in the NH state and 10% in the NL state.

For considering β-catenin activating Mycn, the part of the main GRN equations altered to change the interaction to an activation (equation (2.42)) is now given by

$$\begin{array}{cccc} \vdots & \vdots & \vdots & \vdots \\ \frac{d[M]}{dt} = n \cdot \left(\frac{s_7 \cdot k}{k+[M]} + \frac{s_8 \cdot k}{k+[N]} + \frac{s_9 \cdot [B]}{k_m+[B]} - d_M \cdot [M] \right) + \varphi(0,\sigma_M) \cdot [M] & & & \\ \vdots & \vdots & \vdots & \vdots \end{array} \tag{2.56}$$

showing that the s_9 Hill term has been changed to an activation. Parameter values are in Table 2.6. This interaction is shown schematically in Figure 2.41a. Figure 2.41b shows a 2-parameter continuation on PD and Chiron, where the area above the curve again denotes the values of PD and Chiron for which the system is monostable (NH only). Comparing to Figure 2.29 (Original, red curve) and Figure 2.33b (Original, blue curve) again shows no difference.

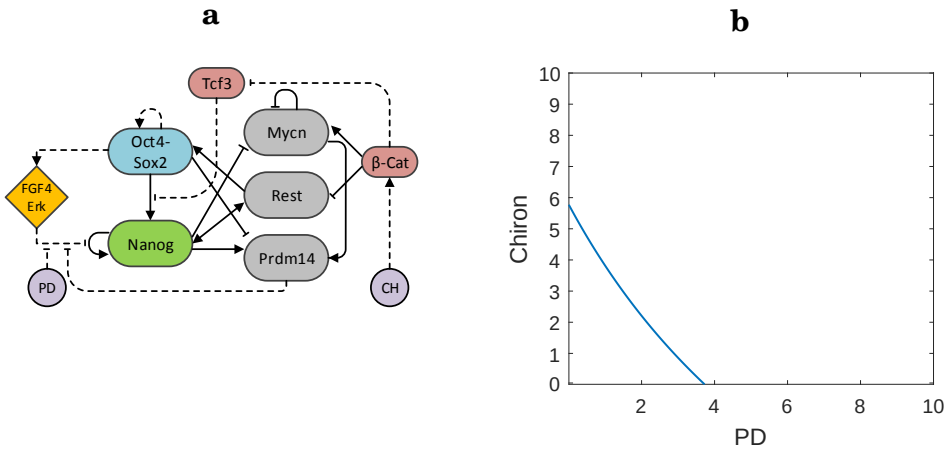


Figure 2.41: β-catenin interaction on Mycn changed from inhibition to activation. **a** Extended GRN of Figure 2.18 with interaction of β-catenin on Mycn changed to activation. **b** 2-parameter continuation of the Saddle-node from Nanog's continuation against PD and Chiron. Above the curve the system is monostable.

Continuations of Nanog's steady-state against its auto-regulatory rate s_4 and simulations of the SDEs show no effect from the change of this interaction to an activation. Figure 2.42a and 2.43a look the same as to the original network shown in Figure 2.19a and 2.24a (for Serum+LIF and 2i+LIF respectively) and Figure 2.42b and 2.43b show the same distributions of Nanog at steady-state in Serum+LIF and 2i+LIF (compare to Figures 2.19b and 2.24b).

Therefore this interaction has been demonstrated to be unimportant for the pluripotency of the simulated system, and was kept as originally found in the database of [168].

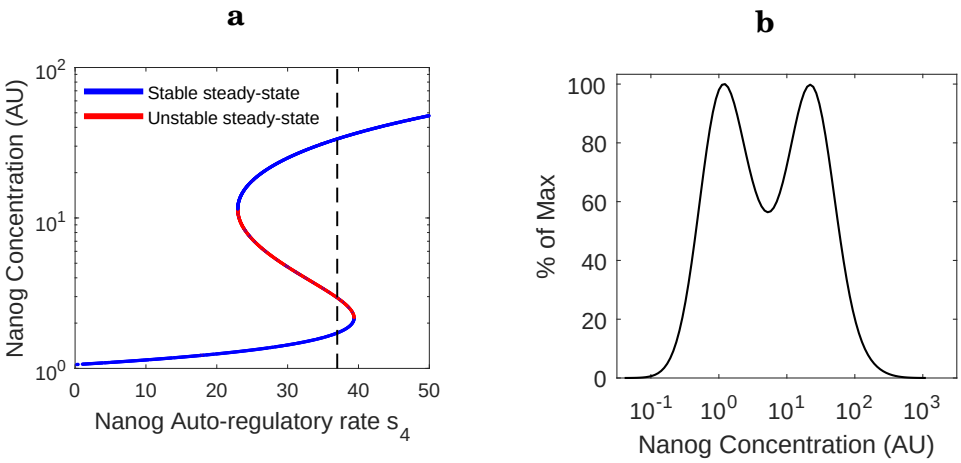


Figure 2.42: β-catenin interaction on Mycn changed to activation results for Serum+LIF parameters. **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, indicating bistability. **b** Distribution of Nanog's steady-state concentration, showing approximately 50% of the cells in the NH state and 50% in the NL state.

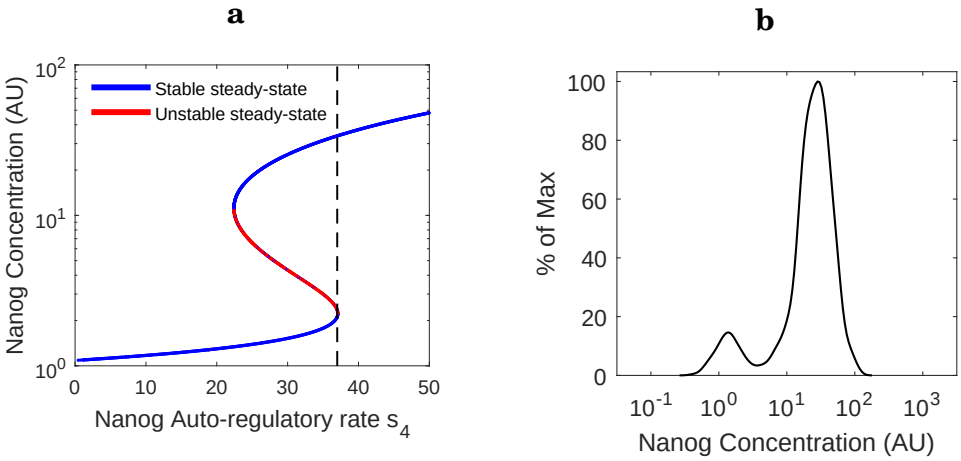


Figure 2.43: β-catenin interaction on Mycn changed to activation results for 2i+LIF parameters. **a** Continuation of Nanog steady-state in the deterministic system (no noise). Blue lines are stable steady-states, red line is unstable. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, though has shifted nearer to the Saddle-node bifurcation, and to the NH monostable regime. **b** Distribution of Nanog's steady-state concentration, showing approximately 90% of the cells in the NH state and 10% in the NL state.

2.4.5 Alternative simpler networks capable of reproducing bistability in Nanog

The deterministic (no noise) bistability of Nanog is caused by an activation feedback (positive auto-regulatory loop), as shown in Section 2.2.5. Reduced networks which keep a positive feedback loop can still recapitulate bistability, even if not directly Nanog activating itself. We wanted to explore which are the key interactions and feedbacks that cause bistability by looking at three different simplified networks with different positive feedback loops established within them. Only continuations of Nanog's steady-state were performed as they are sufficient to show that the respective system is capable of bistable behaviour. The first simplified network we considered is illustrated in Figure 2.44, comprised of the core factors Nanog, Oct4 and Sox2 with an Fgf4/Erk module and β-catenin Tcf3 regulations along with the two inhibitors PD and Chiron.

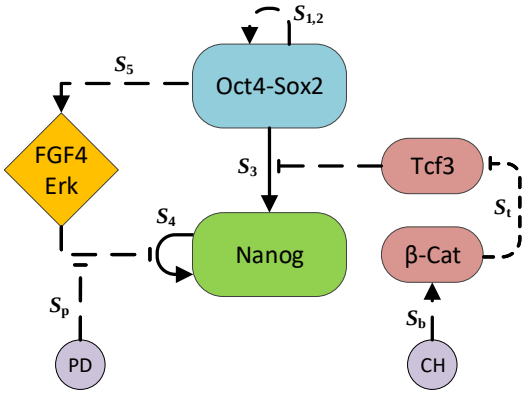


Figure 2.44: Simplified GRN retaining Nanog auto-activation. Solid lines are transcriptional interactions, dashed lines are non-transcriptional interactions. Arrows are activations, flat headed arrows are inhibitions. Core network is made of Nanog, Oct4 and Sox2. FGF4/Erk, Tcf3 and β-catenin are added as targets of the two inhibitors present in 2i+LIF: PD and Chiron (CH). Parameters correspond to equations (2.57) to (2.62).

The equations for the first simplified case of Figure 2.44 are given by

$$p = \frac{s_p \cdot k_p}{k_p + [PD]}, \tag{2.57}$$

$$\frac{d[OS]}{dt} = n \cdot \left(\frac{s_{1,2} \cdot [OS]^2}{(k + [OS])^2 \cdot dO \cdot dS} - d_{OS} \cdot [OS] \right), \tag{2.58}$$

$$\frac{d[N]}{dt} = n \cdot \left(\frac{s_4 \cdot [N]^2}{k^2 + [N]^2 + p \cdot [E]} + \frac{s_3 \cdot [OS]}{k + [OS] + [TC]} - d_N \cdot [N] \right), \tag{2.59}$$

$$\frac{d[E]}{dt} = n \cdot \left(\frac{s_5 \cdot [OS]}{k + [OS]} - d_E \cdot [E] \right), \tag{2.60}$$

$$\frac{d[B]}{dt} = n \cdot \left(\alpha_1 - d_B \cdot \left(\frac{s_b \cdot k_b}{k_b + [CH]} \right) \cdot [B] \right), \tag{2.61}$$

$$\frac{d[TC]}{dt} = n \cdot \left(\alpha_2 - d_{TC} \cdot \left(1 + \frac{s_t \cdot [B]}{k_t + [B]} \right) \cdot [TC] \right), \tag{2.62}$$

N=Nanog; OS=Oct4-Sox2; E=Fgf4/Erk; B=β-catenin; TC=Tcf3. Maximum transcription rates are denoted by S_i, Michaelis–Menten constants by k_i, degradation rates by d_i, constant transcription rates by α_i and a time normalization factor of n.

Parameter values for this system can be found in Table 2.7.

s_p	35*	α_2	40*	d_{PR}	1#
k_p	25	k	10	d_R	1#
$s_{1,2}$	75*	k_t	10	d_E	1#
s_3	2*	k_b	1	d_B	1#
s_4	37*	d_O	1#	d_{TC}	1#
s_5	2*	d_S	1#	n	0.005
s_b	50*	d_{OS}	1#		
s_t	20*	d_N	1#		
α_1	1*	d_m	1#		

(* molecules/minute; # 1/minute)

Table 2.7: Table showing parameter values for GRN of Figure 2.44, described by equations (2.57) to (2.62).

Continuations of the steady-state of Nanog against its auto-regulatory rate s_4 can be seen in Figure 2.45. When simulating the Serum+LIF scenario (PD, CH=0) as shown in Figure 2.45a, the current value of s_4 (black dashed line) intersects through the NL steady-state, the NH steady-state and the unstable steady-state separating them. For 2i+LIF (PD, CH=2) the continuation curve in Figure 2.45b has shifted such that the current value of s_4 (black dashed line) is right on top of the Saddle-node bifurcation. This suggests that Nanog auto-activation is sufficient to

explain the bistable behaviour of Nanog in mESCs, and inclusion of inhibitors is able to alter the stability of the system.

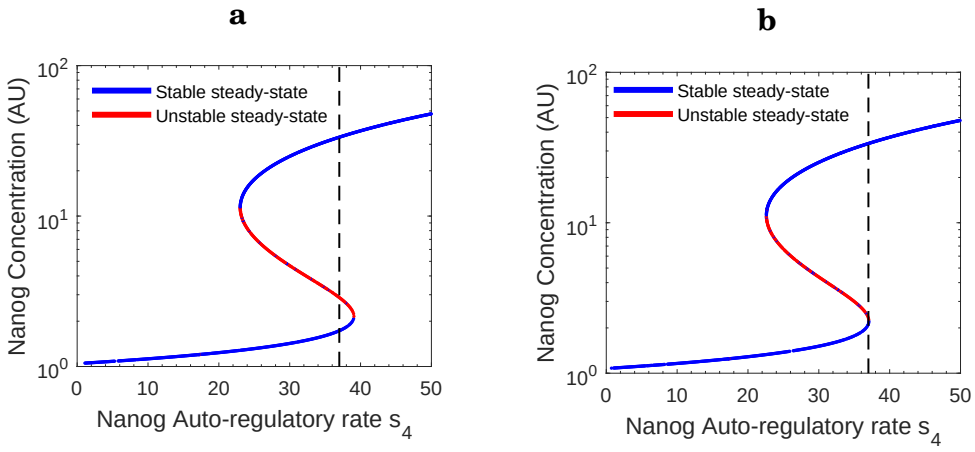


Figure 2.45: Continuation results for GRN of Figure 2.44. Continuation is of Nanog steady-state; blue lines are stable steady-states, red line is unstable. **a** Serum+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, indicating bistability. **b** 2i+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, though has shifted nearer to the Saddle-node bifurcation, and to the NH monostable regime.

The NIA bank analysis we performed when building the extended network showed Nanog to be a direct inhibitor of its own transcription (see Table 2.3). This result is also reported in two independent studies [45, 102]. We therefore explored whether other interactions present in the network could give rise to positive feedback loops capable of reproducing bistability in the steady-state of Nanog, even when changing Nanog's auto-activation into an auto-inhibition. Figure 2.46 shows a GRN for the core factors, with Nanog's auto-activation changed to an inhibition. We additionally added an extra interaction of Nanog activating β-catenin, as in Figure 2.33a. An indirect positive feedback loop is established through β-catenin and Tcf3.

The equations for this system are shown in (2.63) to (2.68) with parameter values in Table 2.8;

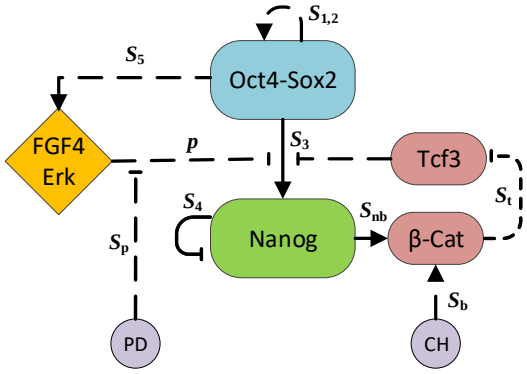


Figure 2.46: Simplified GRN with Nanog auto-inhibition and activation of β-catenin by Nanog. Solid lines are transcriptional interactions, dashed lines are non-transcriptional interactions. Arrows are activations, flat headed arrows are inhibitions. Core network is made of Nanog, Oct4 and Sox2. FGF4/Erk, Tcf3 and β-catenin are added as targets of the two inhibitors present in 2i+LIF: PD and Chiron (CH). Parameters correspond to equations (2.63) to (2.68).

$$p = \frac{s_p \cdot k_p}{k_p + [PD]}, \tag{2.63}$$

$$\frac{d[OS]}{dt} = n \cdot \left(\frac{s_{1,2} \cdot [OS]^2}{(k + [OS])^2 \cdot dO \cdot dS} - d_{OS} \cdot [OS] \right), \tag{2.64}$$

$$\frac{d[N]}{dt} = n \cdot \left(\frac{s_4 \cdot k}{k + [N]} + \frac{s_3 \cdot [OS]}{k + [OS] + [TC]^2 + p \cdot [E]} - d_N \cdot [N] \right), \tag{2.65}$$

$$\frac{d[E]}{dt} = n \cdot \left(\frac{s_5 \cdot [OS]}{k + [OS]} - d_E \cdot [E] \right), \tag{2.66}$$

$$\frac{d[B]}{dt} = n \cdot \left(\alpha_1 + \frac{s_{nb} \cdot [N]^3}{k^3 + [N]^3} - d_B \cdot \left(\frac{s_b \cdot k_b}{k_b + [CH]} \right) \cdot [B] \right), \tag{2.67}$$

$$\frac{d[TC]}{dt} = n \cdot \left(\alpha_2 - d_{TC} \cdot \left(1 + \frac{s_t \cdot [B]^3}{k_t^3 + [B]^3} \right) \cdot [TC] \right), \tag{2.68}$$

N=Nanog; OS=Oct4-Sox2; M=Mycn; PR=Prdm14; R = Rest; E=Fgf4/Erk; B=β-catenin; TC=Tcf3. Maximum transcription rates are denoted by S_i, Michaelis–Menten constants by k_i, degradation rates by d_i, constant transcription rates by α_i and a time normalization factor of n.

Continuations of Nanog's steady-state against its auto-inhibitory rate s_4 can be seen in Figure 2.47. In Serum+LIF (PD, CH=0) the system is bistable, with the black dashed line in Figure 2.47a intersecting two stable steady-states. Increasing the values of the inhibitors for 2i+LIF (PD,

s_p	35*	α_1	1*	d_M	1#
k_p	5	α_2	40	d_{PR}	1#
$s_{1,2}$	75*	k	10	d_R	1#
s_3	35*	k_t	0.15	d_E	1#
s_4	5*	k_b	250	d_B	1#
s_5	2*	d_O	1#	d_{TC}	1#
s_b	50*	d_S	1#	n	0.005
s_t	15*	d_{OS}	1#		
s_{nb}	10*	d_N	1#		

(* molecules/minute; # 1/minute)

Table 2.8: Table showing parameter values for GRN of Figure 2.46, described by equations (2.63) to (2.68).

CH=2) shifts the system further towards the NH state, as in 2.47b. For this system to exhibit bistability the non linearity of the system had to be increased (Hill coefficients of 3 for both s_{nb} and s_t Hill terms) and the sensitivity to changing levels of β-catenin had to be increased by having a low Michaelis–Menten constant for its respective Hill term (k_t in s_t Hill term). This shows that even when considering Nanog's auto-regulation as an inhibition, an indirect positive feedback loop through β-catenin is able to produce bistability.

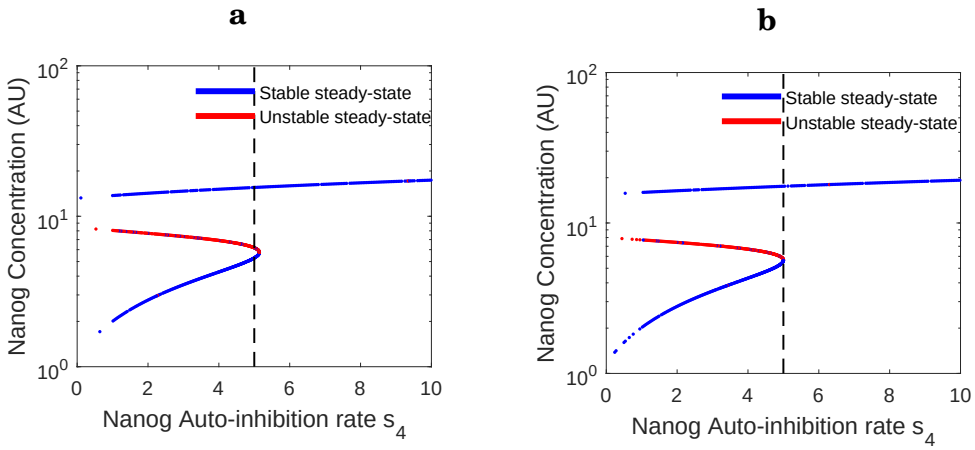


Figure 2.47: Continuation results for GRN of Figure 2.46. Continuation is of Nanog steady-state; blue lines are stable steady-states, red line is unstable. **a** Serum+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, indicating bistbility. **b** 2i+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, though has shifted nearer to the Saddle-node bifurcation, and to the NH monostable regime.

The final simplified network we looked at also took Nanog's auto regulation as an inhibition but neglected Nanog's activation of β-catenin. A different positive feedback loop is established through an interaction present in the extended network: Nanog activating Rest, reciprocally

activating Nanog back. The simplified network considered can be seen in Figure 2.48, described by the equations in (2.69) to (2.75).

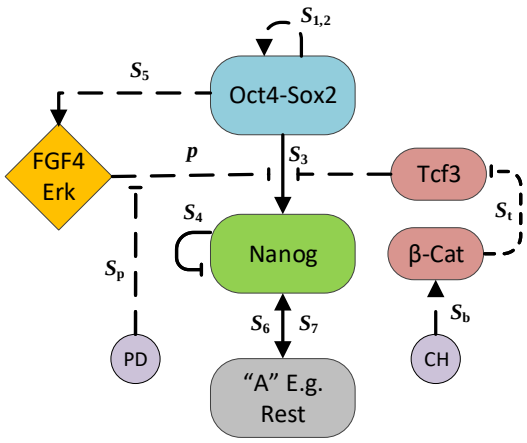


Figure 2.48: Simplified GRN with Nanog auto-inhibition and activation of an additional gene ("A", E.g. Rest) by Nanog, which in turn activates Nanog back. Solid lines are transcriptional interactions, dashed lines are non-transcriptional interactions. Arrows are activations, flat headed arrows are inhibitions. Core network is made of Nanog, Oct4 and Sox2. FGF4/Erk, Tcf3 and β-catenin are added as targets of the two inhibitors present in 2i+LIF: PD and Chiron (CH). Parameters correspond to equations (2.69) to (2.75).

The interactions of this network are modelled by:

$$p = \frac{s_p \cdot k_p}{k_p + [PD]}, \tag{2.69}$$

$$\frac{d[OS]}{dt} = n \cdot \left(\frac{s_{1,2} \cdot [OS]^2}{(k + [OS])^2 \cdot dO \cdot dS} - d_{OS} \cdot [OS] \right), \tag{2.70}$$

$$\frac{d[N]}{dt} = n \cdot \left(\frac{s_4 \cdot k}{k + [N]} + \frac{s_3 \cdot [OS]}{k + [OS] + [TC] + p \cdot [E]} + \frac{s_6 \cdot [R]^2}{{k_n}^2 + [R]^2} - d_N \cdot [N] \right), \tag{2.71}$$

$$\frac{d[R]}{dt} = n \cdot \left(\frac{s_7 \cdot [N]^2}{k^2 + [N]^2} - d_R \cdot [R] \right), \tag{2.72}$$

$$\frac{d[E]}{dt} = n \cdot \left(\frac{s_5 \cdot [OS]}{k + [OS]} - d_E \cdot [E] \right), \tag{2.73}$$

$$\frac{d[B]}{dt} = n \cdot \left(\alpha_1 - d_B \cdot \left(\frac{s_b \cdot k_b}{k_b + [CH]} \right) \cdot [B] \right), \tag{2.74}$$

$$\frac{d[TC]}{dt} = n \cdot \left(\alpha_2 - d_{TC} \cdot \left(1 + \frac{s_t \cdot [B]}{k_t + [B]} \right) \cdot [TC] \right), \tag{2.75}$$

N=Nanog; OS=Oct4-Sox2; R = Rest; E=Fgf4/Erk; B=β-catenin; TC=Tcf3. Maximum transcription rates are denoted by S_i, Michaelis–Menten constants by k_i, degradation rates by d_i, constant transcription rates by α_i and a time normalization factor of n.

Parameter values for this system can be found in Table 2.9.

s_p	35*	s_t	20*	d_{OS}	1#
k_p	10	α_1	1*	d_N	1#
$s_{1,2}$	75*	α_2	40*	d_M	1#
s_3	6*	k	10	d_{PR}	1#
s_4	1*	k_n	45	d_R	1#
s_5	2*	k_t	6	d_E	1#
s_6	48*	k_b	1	d_B	1#
s_7	50*	d_O	1#	d_{TC}	1#
s_b	50*	d_S	1#	n	0.005

(* molecules/minute; # 1/minute)

Table 2.9: Table showing parameter values for GRN of Figure 2.48, described by equations (2.69) to (2.75).

Continuations of Nanog's steady-state against the activation rate of Rest on Nanog, s_6, are shown in Figure 2.49. In Serum+LIF the deterministic system is bistable, with the chosen value of s_6 (dashed black line) intersecting both the NH and NL steady-states. When PD and Chiron are introduced into the system, the 2i+LIF results show the system shifted further to the NH steady-state with the value of s_6 (dashed black line) much closer to the Saddle-node bifurcation.

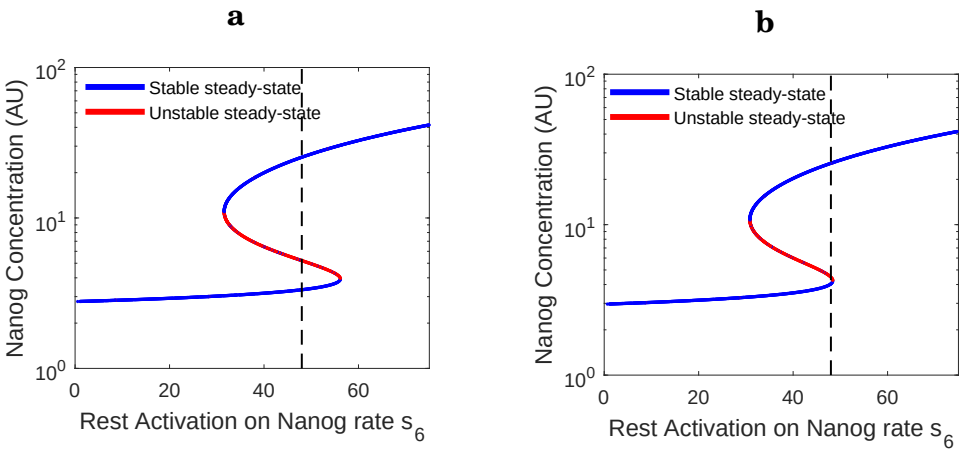


Figure 2.49: Continuation results for GRN of Figure 2.48. Continuation is of Nanog steady-state; blue lines are stable steady-states, red line is unstable. **a** Serum+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, indicating bistbility. **b** 2i+LIF. The current transcription rate s_4 (black dashed line) intersects both the NH and NL steady-state, though has shifted nearer to the Saddle-node bifurcation, and to the NH monostable regime.

2.4.6 Agent-based model

All analysis in this section was performed by the following collaborators: Daniel Ward and Martin Homer from the Department of Engineering Mathematics in Bristol; Elisa Pedone from Engineering Mathematics and the school of Cellular and Molecular Medicine in Bristol; Alexander G. Fletcher from the School of Mathematics and Statistics and the Bateson centre in the University of Sheffield; and Lucia Marucci from the Department of Engineering Mathematics, the School of Cellular and Molecular Medicine and BrisSynBio of the University of Bristol.

All the previous modelling and analysis ignored physical cell to cell interactions and the cell-cycle. The cell-cycle is composed of different phases, namely G1, S and G2, through which a cell progresses before dividing. More succinctly, the length of the cell-cycle is the time it takes for a cell to divide. Three studies disagreed in their results of how 2i+LIF and Serum+LIF affect the cell-cycle. One [26] reported a quicker cell-cycle time in 2i+LIF, another [71] reported a slower cell-cycle time in 2i+LIF, and a third [145] reported no difference between the two media. We therefore performed our own culturing of mESCs in Serum+LIF and 2i+LIF, measuring how the rate of proliferation varies between the two media. We found a significantly slower proliferation rate in 2i+LIF as compared to Serum+LIF, as shown in Figure 2.50.

An agent based model was developed to match the progression of the cell-cycle within the two media. The was created in the Chaste modelling framework [93, 110], an extensive library

in C++ which contains tools for agent based modelling of cell/tissue cultures. Each cell agent contains its own realization of the GRN, upon which gene concentration levels within affect cell-cycle and cell states. The cell-cycle time, derived from the GRN, is then fed into the model that describes physical interactions between neighbouring cells (see Figure 2.55 for snapshot of physical agent system). Two different genes within the GRN are candidates from which the cell-cycle can be based on, Nanog and Mycn. Nanog levels have been shown to correlate positively with proliferation rate (at least in Serum+LIF) [61]. The Myc family of TFs (includes Myc and Mycn) has been shown to positively regulate cell growth proliferation in various cell types [30]. When using Nanog to define a mESCs cell-cycle time, cells are classified as either Nanog high (NH) or Nanog low (NL), using a threshold level to separate them. The same approach is used when using Mycn for the cell-cycle, except mESCs are classified as Mycn high (MH) or Mycn low (ML). A mESC classified as NH will progress through the cell-cycle faster than for NL and will therefore have a shorter cell-cycle duration, with the same behaviour for MH cells versus ML cells. Distributions of cell-cycle times were created for NH, NL, MH and ML, where the time it takes for an agent to divide would be a variable randomly drawn from the distribution corresponding to its state (e.g. NH state). The cell-cycle distributions were altered to match the proliferation rate of the simulated cell agents with that of the real world mESCs we measured experimentally.

We first used Nanog to couple gene expression and cell-cycle duration as in [61], where simulated mESC agents had their cell-cycle times altered to match the proliferation we measured in Serum+LIF. With the allocated cell-cycle times ($CC_{NH} = 8.5 \pm 1h$ and $CC_{NL} = 13.5 \pm 1h$, where CC_{NH} and CC_{NL} are cell-cycle times for NH and NL cells respectively), the simulated Serum+LIF proliferation rate had a root-mean squared error (RMSE) of 2.35 compared to the data, as shown in Figure 2.50a (blue line). Distributions of Nanog concentration in the population of cells showed approximately 50% in NL and 50% in NH, consistent with the SDE model and experimental data (compare Figure 2.51a to 2.19b and 2.2b). Using the cell-cycle times fitted to Serum+LIF but with 2i+LIF parameters (PD=2, CH=2, Nanog noise reduced 20%) resulted in a distribution of Nanog concentrations consistent with the SDE model and experimental data of approximately 90% NH and 10% NL (compare Figure 2.51b to 2.24b and 2.2b). However, the proliferation rate of mESCs in the agent model was increased in 2i+LIF, in direct contrast to experiment (red line in Figure 2.50a) with a RMSE of 22.59.

We then fitted cell-cycle times to the 2i+LIF experimental rates by increasing the cell-cycle times ($CC_{NH} = 13.175 \pm 1h$ and $CC_{NL} = 20.925 \pm 1h$) resulting in a RMSE of 1.07 between the model and 2i+LIF data shown in Figure 2.50b (red line). However these times were poor at predicting the growth rate in Serum+LIF (blue line in Figure 2.50b) with a RMSE of 11.01. This shows that coupling the cell-cycle to Nanog is unable to reproduce the experimental proliferation rates observed between the two media.

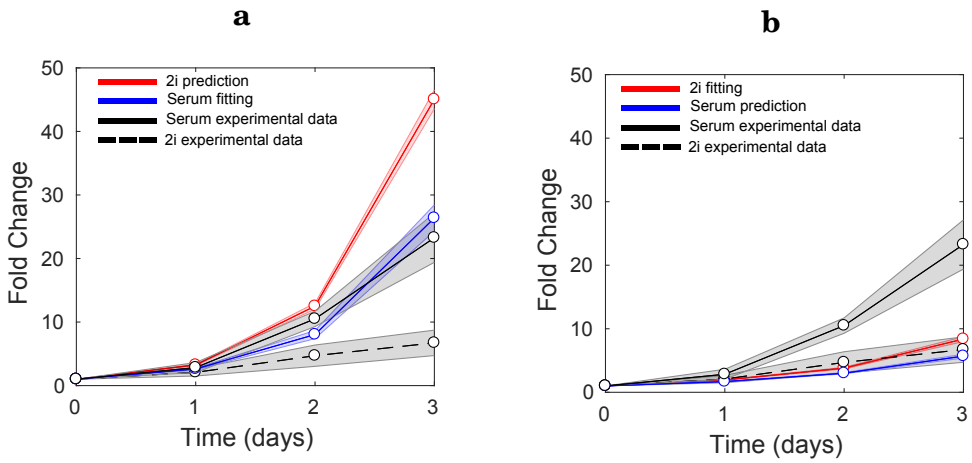


Figure 2.50: Comparison of experimental and model simulated fold change in proliferation dynamics of mESCs in the two different media. In the model, the cell-cycle is coupled to Nanog levels. **a** Serum+LIF data were fitted, and 2i+LIF behaviour predicted by the model. **b** 2i+LIF data were fitted, and Serum+LIF behaviour predicted by the model.

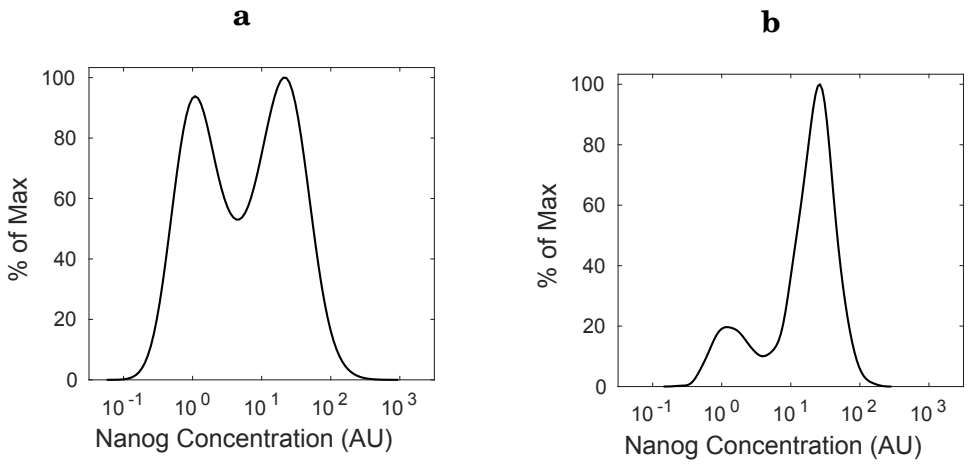


Figure 2.51: Nanog steady-state distribution, with cell-cycle coupled to Nanog concentration. **a** Serum+LIF parameters; approximately 50% of the mESCs are in a NH state and 50% in the NL. **b** 2i+LIF; approximately 90% of the mESCs are in a NH state and 10% in the NL.

The same procedure was carried out again, coupling cell-cycle to Mycn instead of Nanog. Proliferation was fitted to the measured rate in Serum+LIF, with the cell-cycle time distributions ($CC_{MH} = 8.5 \pm 1h$ and $CC_{ML} = 15 \pm 1h$, where CC_{MH} and CC_{ML} are cell-cycle times for MH and ML cells respectively) giving a RMSE of 2.02. The distribution of Nanog concentration within the cell population again matched the SDE model of around 50% NH and 50% NL cells (compare Figure 2.53a to 2.19b and 2.2b). Using these cell-cycle times, but simulating 2i+LIF gave a distribution of Nanog matching the SDE model and experimental data of approximately

90% NH, 10% NL (Compare Figure 2.53b to 2.24b and 2.2b). The proliferation profile using the Serum+LIF fitted cell-cycle times with 2i+LIF medium parameters gave a much better fit to the measured rate in 2i+LIF (red line in Figure 2.52a) with a RMSE of 3.08 (compared to 22.59 with the Nanog cell-cycle). The growth rate was then fitted to the 2i+LIF proliferation data with the fitted cell-cycle times ($CC_{MH} = 9.5 \pm 1h$ and $CC_{ML} = 16 \pm 1h$) resulting in a RMSE of 1.33 (red line in Figure 2.52b). Using these times with Serum+LIF parameters gave a better fit than before (RMSE of 4.41 compared to 11.01), and the qualitatively correct change in proliferation of mESCs dividing faster in Serum+LIF than in 2i+LIF (blue line in Figure 2.52b).

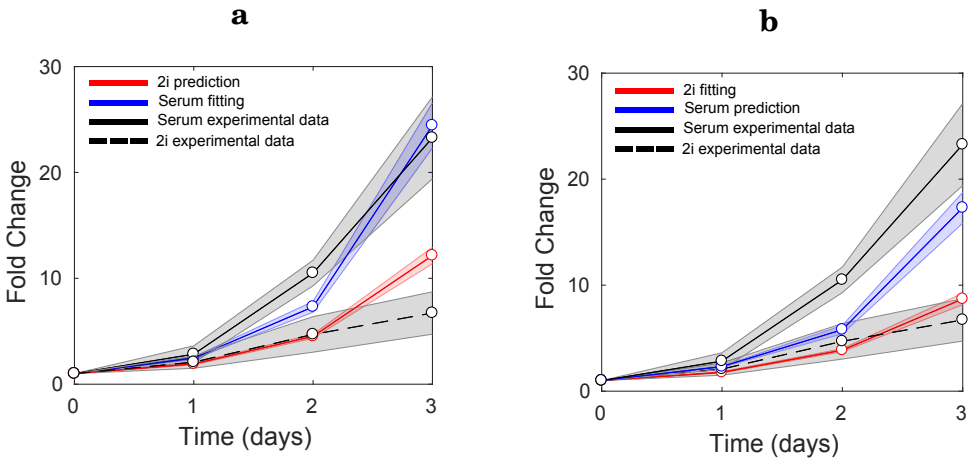


Figure 2.52: Comparison of experimental and model simulated fold change in proliferation dynamics of mESCs in the two different media. In the model, the cell-cycle is coupled to Mycn levels. **a** Serum+LIF data were fitted, and 2i+LIF behaviour predicted by the model. **b** 2i+LIF data were fitted, and Serum+LIF behaviour predicted by the model.

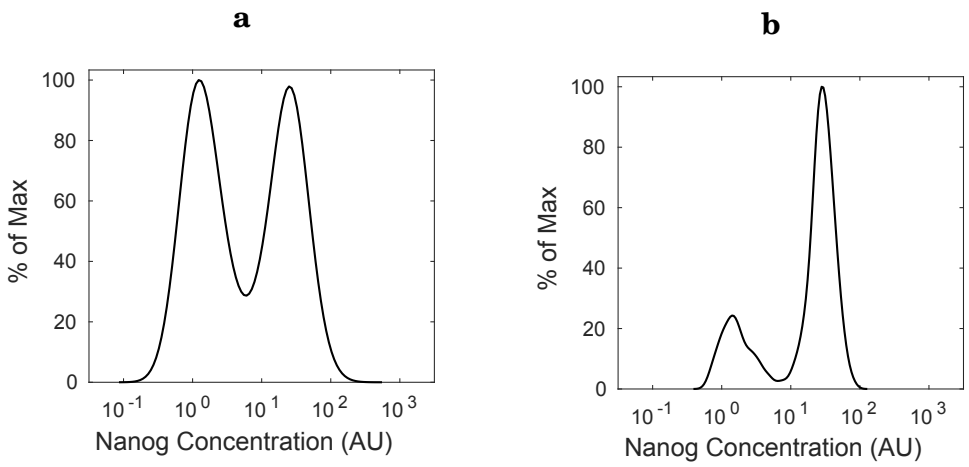


Figure 2.53: Nanog steady-state distribution, with cell-cycle coupled to Mycn concentration. **a** Serum+LIF; approximately 50% of the mESCs are in a NH state and 50% in the NL. **b** 2i+LIF; approximately 90% of the mESCs are in a NH state and 10% in the NL.

Sorting experiments were simulated with the agent model for the Mycn dependant cell-cycle model, as shown in Figure 2.54. The agent based model shows the same cell dynamics upon sorting as the SDE model, comparing Figure 2.54a to 2.20a for Serum+LIF and 2.54b to 2.25b for 2i+LIF.

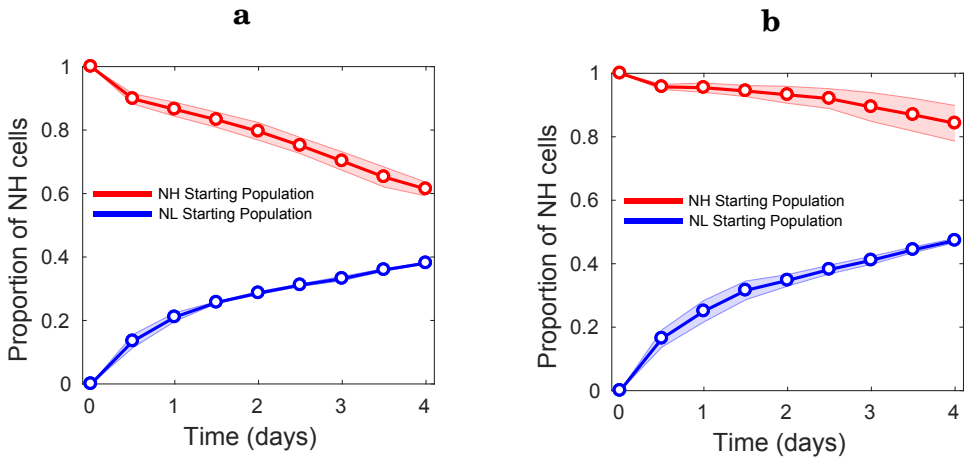


Figure 2.54: Agents sorted by Nanog concentration into NH and NL, re-establishing steady-state concentrations (Mycn dependent cell-cycle). Shaded areas represent standard error. **a** Sorting with Serum+LIF parameters, with agents tending towards a population with approximately 50% NH and 50% NL. **b** Sorting with 2i+LIF parameters, agents tend towards approximately 90% NH and 10% NL.

Snapshot of agents used to model mESCs can be seen in Figure 2.55. Starting agent configurations are in Figures 2.55a and 2.55c for Nanog and Mycn respectively, and snapshots after

a few cell divisions are shown in Figures 2.55b and 2.55d for Nanog and Mycn concentrations respectively.

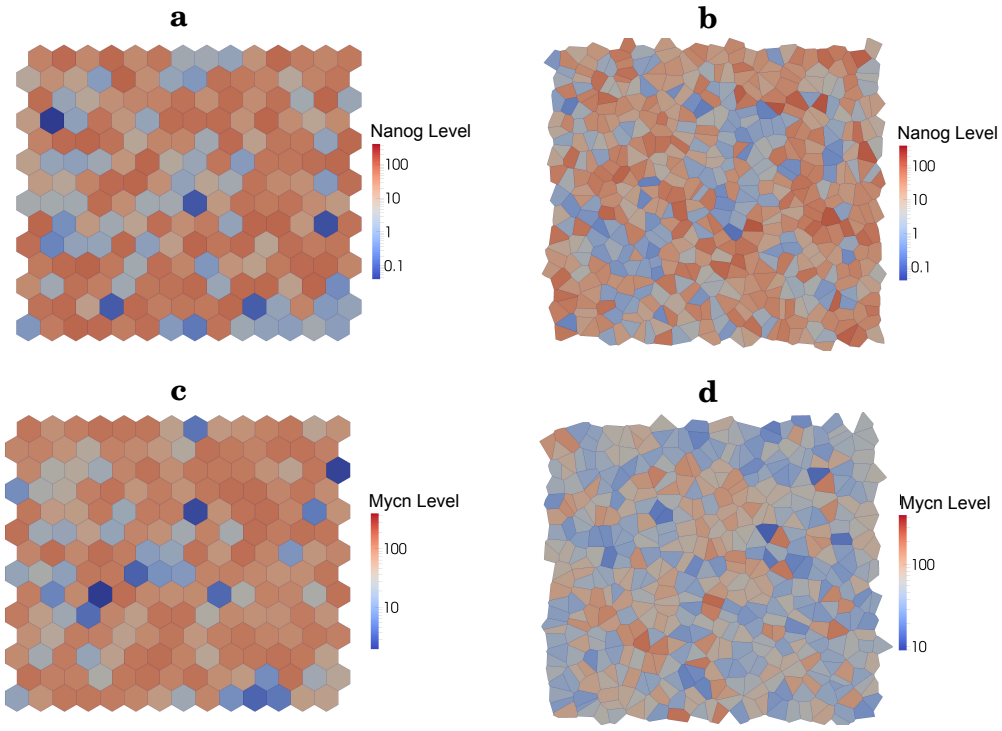


Figure 2.55: Snapshot of cell agents with colour representing gene concentration. **a**,**c** Start of simulation for Nanog and Mycn respectively. **b**,**d** Mid way through simulation for Nanog and Mycn respectively.

2.5 Discussion

We derived a new GRN, describing the interactions between genes governing mESC pluripotency behaviour. We performed both deterministic and stochastic simulations of our GRN, and were able to reproduce experimental data. Furthermore, an agent based extension to the GRN was developed to account for mESC cell-cycle.

Our network reproduced distributions of the key pluripotency gene Nanog, and how the distribution changes when cultured in serum+LIF compared to in 2i+LIF. We correctly reproduce ND cell distributions with our model, showing greater heterogeneity in Serum+LIF, and a persistent sub-population of NL cells in 2i+LIF. The GRN reproduces sorting data as obtained from flow cytometry sorting and individual cell time courses of Nanog match observed time courses, stochastically fluctuating between high and low states within a single cell-cycle. Our numerical continuation analysis on PD and Chiron show that both contribute to the regulation of

the stability and dynamics of the system. Furthermore, our gene deletion simulations correctly predict a reduction in pluripotency upon Prdm14 and increase in pluripotency for Tcf3 deletion.

We demonstrated via the simpler topologies results that the newly added genes to the network, Mycn, Prdm14 and Rest, were not necessary to reproduce Nanog dynamics. However, these newly added genes were each found to have a key role within the GRN. Prdm14 maintains the correct proportion of NH and NL mESCs via its interaction through the FGF4/Erk pathway, as shown with the Prdm14 deletion results. Rest is a candidate that can supply a positive feedback on Nanog, even in the case of Nanog's direct regulation being an inhibition, as shown by the reduced network results. The agent based model results indicate Mycn as a key gene regulating mESC proliferation, as we show that it was not possible to account for mESC proliferation dynamics when coupling Nanog to the cell-cycle, as suggested by a previous model [61].

The two main features of our GRN giving rise to a heterogeneous distribution of Nanog are; a positive feedback loop on Nanog giving it deterministic bistability, and noise which allows individual cells to switch between the two steady-states. We show there is an interplay between noise and determinism, and that the system can be very robust to changes in the deterministic model. Many of our results show changes to the deterministic model make little observable difference to the full SDE realization with those parameters.

Experimental evidence, and our own NIA analysis, suggests that Nanog's direct auto regulation is an inhibition not an activation. We show that other indirect positive feedback loops can be generated within our GRN, even in the case of considering Nanog's direct regulation as an inhibition. If these indirect positive feedbacks are not considered, another alternative mechanism that could explain the heterogeneity of Nanog, would be a negative feedback on Nanog giving rise to a limit cycle alternating between the two steady-states. However, a published model incorporating a negative feedback loop on Nanog resulting in a limit cycle [52] was not consistent with recent experimental results based on mESC single-cell resolution data [139]. Therefore, with the current evidence, a positive feedback with added transcriptional noise is the better modelling approach.

The main limitation of our model is the complexity of these GRNs; there are thousands of genes and molecular substrates all interacting with each other and we only look at a handful of factors. Another limitation is that many of the parameters of the system were not fitted to data, only qualitatively reproducing behaviour, and could be considered arbitrary. Even with these limitations, our analysis still provides useful insights into the stability and dynamic behaviour of the GRN governing pluripotency in mESCs.

CHAPTER 3

Background of fMRI Analysis: A Systematic Review of Resting-State Studies in Multiple Sclerosis

This chapter gives the background and motivation for the work carried out in Chapter 4 in the form of a systematic review. A systematic review is a type of literature review generally used for clinical practise, in which one follows explicit and transparent methods in order to minimise bias. The process was written up prior to the review in a protocol which is publicly published [54]. The review follows a standard set of stages set out in the protocol, and as such it is repeatable and transparent. Two reviewers were mainly involved in this review; myself as main reviewer and Christelle van Anterwerpen of CRICBristol as second reviewer.

3.1 Introduction

Multiple Sclerosis (MS) is characterised as an inflammatory demyelinating disease of the central nervous system (CNS), which results in widespread plaques in the brain and spinal cord and results in demyelination of the axons. MS is a common neurological disorder and leading cause of disability in young adults. It has considerable negative impact on quality of life and its prevalence around the world is increasing [17]. MS represents a significant health burden both financially and socially. Mackenzie *et al.* 2013 [80] estimated that the cost of MS in the UK in 2010 alone was approximately 3.8 billion euros. In 2010 the prevalence of MS was 258.8 per 100,000 in women and 113.3 per 100,000 in men [80]. The pathology of MS is diverse with both physical and neuropsychiatric symptoms [44]. Two of the most common symptoms are fatigue and cognitive dysfunction, with over 92% of patients reporting fatigue as one of the most debilitating effects of the disease [16], and 45-65% presenting with cognitive deficits on clinical assessments [34].

Early studies suggested that cognitive impairments were directly related to neurological changes [144]. However more recent research suggests only a weak relationship between the changes in the brain and cognitive impairments in MS [81]. One possible explanation for these results may be due to neuroplasticity, i.e., the ability to reorganise and form new connections in the brain. Neuroplasticity may allow individuals with MS to maintain normal performance despite the accumulating tissue damage. Understanding the mechanisms as to how the brain compensates and reacts to this damage is an important first step for future clinical applications. Understanding the neuroplasticity may provide insights into the neural mechanisms in MS, and may even lead to an objective measurement to determine the effectiveness of treatments or therapies.

MS is comprised of several phenotypes. Approximately 80% of patients manifest a relapse-remitting course, relapse-remitting MS (RRMS), which after several years can be succeeded by a secondary progressive phase, secondary progressive MS (SPMS). In a minority of patients the relapse-remitting stage does not manifest and the progressive stage occurs at onset, primary progressive MS (PPMS) [15, 74]. Early studies of MS suggested that it was a disease related to white matter atrophy and that grey matter lesions were rarely seen [117]. More recent studies have established that brain atrophy affects both the white matter and grey matter in MS [73, 111, 121, 169]. However, structural imaging studies are not able to investigate how this damage affects the function of the brain. An indirect measure of brain activity is functional magnetic resonance imaging (fMRI). This is a non-invasive method of measuring brain activity, focusing on blood flow through the brain via the blood oxygen level dependent (BOLD) signal. The BOLD signal of a region is an indirect measure of neuronal activity in that region, as a more active region of the brain requires more oxygenated blood. fMRI is most commonly used to see which regions of the brain are activated by a certain task or stimuli. In MS specifically, multiple studies have used fMRI to investigate cognitive impairment in domains such as memory and attention [14, 81, 123, 156] as well as establish possible neural correlates of fatigue [82]. Recently, there has been a growing interest into applying fMRI at rest, termed resting-state fMRI (rs-fRMI) [119]. This technique measures spontaneous fluctuations of the BOLD signal and an associated measure of the baseline functional organisation of the brain called functional connectivity (FC). Changes in measured FC between MS and healthy controls (HC) can be used as evidence for neuroplasticity in MS.

The main focus of this review was to investigate the evidence of neuroplasticity in MS from rs-fMRI data. Further aims were to evaluate the different data analysis techniques and methods applied to rs-fMRI, to assess the advantages and disadvantages of the individual analysis techniques and to understand the implications of MS on FC and neuroplasticity.

3.2 Materials and methods

We searched Embase, Medline, PsychINFO and Scopus from the first year of successful fMRI (1991) up until the start date of the review (21 March 2017). Search terms were grouped into four categories: MS terms, fMRI terms, resting-state terms and neuroplasticity terms. Multiple words were searched for in each category and the union of studies found under a specific term formed a group. See Table 3.1 for the search words used. Articles of interest were in the intersection of all four of these groups as shown in Figure 3.1.

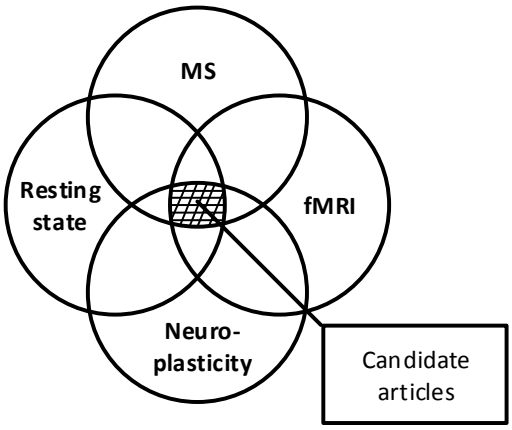


Figure 3.1: Schematic of search term groups. Terms related to a group were searched for, and the union of results formed the respective group.

MS terms	fMRI terms	Resting-state terms	Neuroplasticity terms
ms	fmri	resting state	neuroplasticity
multiple sclerosis	f mri	functional connectivity	neuro plasticity
rrms	functional mri	functional connectivity network	neural plasticity
rr ms	functional mr imaging	resting state network	brain plasticity
relapsing remitting ms	functional magnetic resonance imaging	resting state functional connectivity	neuro reorganisation
relapsing remitting multiple sclerosis	bold	resting state connectivity	neural reorganisation
spms	blood oxygen level dependent	resting state fmri	brain reorganisation
sp ms		rs fmri	neuro reorganization
secondary progressive ms		rsfmri	neural reorganization
secondary progressive multiple sclerosis		default mode network	brain reorganization
ppms		dmn	neuro compensation
pp ms		rsn	neural compensation
primary progressive ms			brain compensation
primary progressive multiple sclerosis			neuro compensatory
cis			neural compensatory
clinically isolated syndrome			brain compensatory

Table 3.1: Search terms used, in their respective term group.

Inclusion criteria were: resting-state fMRI studies of MS patients, where MS is defined as per National Institute for Health and Care Excellence (NICE) guidelines [2], Centers for Disease

Control and Prevention (CDC) criteria [108], World Health Organisation [17] or the McDonald criteria [89, 115, 116]. Additional inclusion criteria were resting-state fMRI studies of MS alone, or MS in comparison to healthy or other clinical populations. Exclusion criteria were: studies older than the first published paper defining the BOLD mechanism in 1991 and after the start date of review, 21 March 2017, and task-based fMRI, i.e. studies with no pure resting-state acquisition.

Two reviewers (SG, CVA) independently assessed the articles of interest for eligibility using the tool Covidence [1], where disagreements were resolved via discussion and consensus. We assessed the quality of studies in terms of risk of bias using seven criteria (Table 3.9): research objective, recruitment procedure, inclusion/exclusion criteria, population demographics, imaging protocol, comparison group and preprocessing protocol [105].

3.3 Results

We found 106 candidate articles from searching the mentioned databases.

Any review articles among the 106 articles were read and any papers they cited relating to MS, resting-state and fMRI were manually included into the title and abstract screening. We found 23 additional papers cited in the reviews, resulting in a total of 129 candidate articles. Of these, 60 were excluded at the title and abstract screening stage. Reasons for exclusion were studies of patients without MS, no resting-state, no fMRI (e.g. Diffusion Tensor Imaging (DTI) only) and not original research studies (i.e. reviews). At the full text screening stage, 41 studies were excluded. Reasons for exclusion were repeated papers, studies not in English, no diagnostic criteria given, abstract only for poster/oral presentation, and no pure resting-state (i.e. time slices from different periods of rest in-between tasks, concatenated together). In total, of the 129 articles 28 met the inclusion criteria. See Figure 3.2 for a breakdown of the article numbers in each stage of the review.

3.3.1 Quality assessment

The number of studies found investigating pure resting-state fMRI is low, however (in terms of how much a study fulfilled our criteria) the quality of these studies is high. Only four studies did not fulfil all the criteria; see Table 3.9 for quality assessment results.

When MS participants were compared to HC participants the two groups were well matched in nearly all the studies, the majority matching for age, sex and education level. This is good study design as these variables have been shown to affect FC [32], and not matching them could lead to erroneous results.

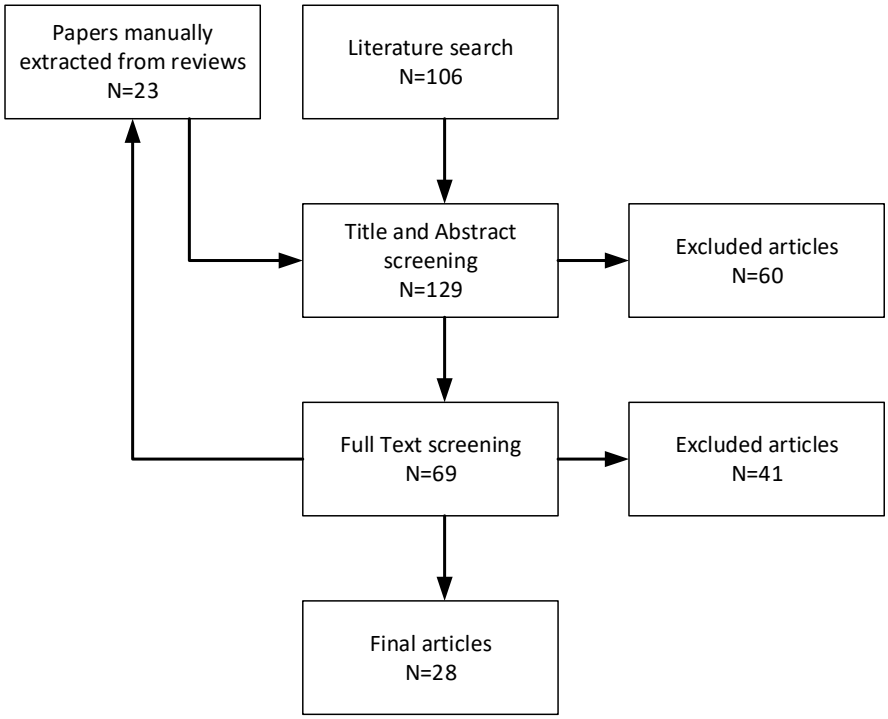


Figure 3.2: Diagram of review steps.

3.3.2 MS phenotypes and diagnostic criteria

All the articles used either the McDonald criteria [89], 2005 revisions to the McDonald critera [116] or 2010 revisions to the McDonald criteria [115].

The phenotype of MS in participants was Relapsing Remitting MS (RRMS) in 19 studies; of which two [146, 157] used the original McDonald criteria, nine [12, 13, 50, 68, 75, 76, 114, 122, 171] used the 2005 revision to the McDonald criteria and eight [28, 29, 31, 41–43, 170, 172] the 2010 revision. A combination of RRMS and Clinically Isolated Syndrome (CIS) was used in two studies; one [124] using the 2005 revision and one [51] the 2010 revision. Another study used a combination of RRMS and CIS but also included unspecified MS type participants [60]. This study used the original McDonald criteria. A combination of RRMS and Secondary Progressive MS (SPMS) was included in three studies; two [35, 133] using the 2005 revision and one [65] the 2010 revision. Finally, three studies used a combination of RRMS, Primary Progressive MS (PPMS) and SPMS; two [131, 135] using the 2005 revision and one [132] using the 2010 revision of the McDonald criteria.

3.3.3 Study aims and size

Main study aims varied: for four studies [75, 76, 170, 172] the aim was to determine differences between MS and HC; another study [171] also compared MS and HC though their main aim was exploring how FC measurements were correlated with clinical neurological scales. For two studies the main aim [12, 28] was comparing cognitively preserved with cognitively impaired MS patients. Two studies [13, 31] examined the impact of a computer rehabilitation program. One study [29] focused on the impact of fatigue. Six studies [35, 41, 42, 68, 131, 132] investigated the impact of disability on FC. Four studies [60, 65, 146, 157] focused on cognition, one of which [65] focused on memory performance, and two [146, 157] examined both cognition and disability. Four studies [43, 51, 114, 124] explored changes in FC evidenced in early MS (i.e. CIS or early RRMS). One study [50] explored changes in FC due to optic neurosis. One study [133] focused on how gender affects FC and cognition. One study [135] looked at how pain affected FC, and one unique study [122] described a classifier they developed, based on a random forest scheme, to classify participants' fMRI data into either MS or HC.

The maximum number of MS participants in a study was 157, with the minimum number being 13. For studies that included HC participants the maximum and minimum number of HCs were 50 and 11 respectively. Overall, the number of MS participants in the studies were 40 ± 32 (mean ± std) and for HC was 24 ± 11 (mean ± std) (for studies that included HCs). See Table 3.8 for full study demographic details.

3.3.4 Data acquisition and analysis methods

Two-thirds of the studies [12, 13, 28, 29, 31, 50, 68, 75, 76, 114, 122, 131, 132, 135, 146, 157, 170–172] had scan length times of 9 minutes or less (9 ± 4 minutes, mean ± std); 85% [12, 13, 28, 29, 31, 41–43, 50, 65, 68, 75, 76, 114, 124, 131–133, 135, 146, 157, 170–172] had 300 volumes or less (241 ± 109 volumes, mean ± std); 85% [28, 29, 31, 35, 41–43, 51, 60, 65, 68, 75, 76, 114, 124, 131–133, 135, 146, 157, 170–172] had a time to repetition (TR) of greater than or equal to 2 seconds (2344 ± 653 milliseconds, mean ± std).

Various methods were used to explore neuroplasticity in MS, which we split into six categories: independent component analysis (ICA), seed to voxel, region of interest (ROI), automated anatomical labelling (AAL) network measures, amplitude of low frequency fluctuations (ALFF) and voxel to voxel. Below we describe how each method defines locations within the brain and how they calculate FC.

ICA measures coherent fluctuations between groups of voxels and was used in 13 studies: 10 exclusively [12, 13, 28, 41, 43, 50, 68, 114, 124, 135], one [29] in combination with a seed based analysis, one [35] in combination with a ROI based analysis and one [170] used ICA to

define their own ROI locations. A Seed to voxel technique correlates every voxel in the brains BOLD signal separately with a chosen seed locations (usually averaged) BOLD signal. Groups of voxels with significant correlation values are then tested with multiple comparison and cluster size statistical tests. Voxel clusters passing these tests are then reported to be affected. Six studies were identified as using a seed to voxel analysis: five exclusively [31, 75, 131, 146, 157] and one in combination with ICA [29]. The study which used it in combination with ICA used Pearson's correlation coefficient to measure FC whilst all the other studies did not specify how they calculated it. A ROI analysis method looks at multiple regions of averaged BOLD signal voxels and compares them to each other, rather than comparing at a voxel level. Four studies used a ROI analysis in their papers. Two used the AAL atlas to define ROIs, focusing on thalami ROIs in one study [133] or adding custom hippocampal ROIs to focus on in another [65]. Both these studies measured FC using synchronization likelihood (SL). Another study [170] used ICA to define ROI locations, measuring FC between them using Pearson's correlation coefficient. One study [35] used ICA on a ROI defined by a previous activation study.

The AAL atlas is a brain segmentation based on anatomy. Six studies were identified as using an AAL atlas analysis; two [65, 133] to define ROIs in a ROI analysis, three [42, 51, 132] to measure network properties and one [122] using the AAL atlas to build a classifier. Of the studies that measured network properties one [42] did not specify how they measured FC, one [51] used Pearson's correlation coefficient and the other [132] used SL. The study which built a classifier [122] used Pearson's correlation coefficient to measure FC. Two studies [76, 171] used Amplitude of low Frequency Fluctuations (ALFF) which measures intrinsic oscillation amplitude. Two studies performed a voxel to voxel analysis (Table 6), one [60] between 40,000 voxels spanning the cortex using Pearson's correlation coefficient, and the other [172] measuring Pearson's correlation coefficient between all the voxels in a sensorimotor network mask.

3.3.5 Neuroplasticity in MS

The studies found in this review demonstrated possible neuroplasticity in MS either in terms of anatomy (i.e. a physical part of the brain), in terms of a resting-state functional network (which can involve many distinct regions of the brain) or as a network measure.

ICA

Thirteen studies used ICA [12, 13, 28, 29, 35, 41, 43, 50, 68, 114, 124, 135, 170]. See Table 3.2 for results.

The default mode network (DMN) was found to be effected in seven of these [12, 13, 28, 43, 124, 135, 170]. Two studies showed increased DMN FC in MS compared to HC [41, 65], one study [12] found both increased and decreased FC in specific regions within the DMN, one [28] study showed a decrease in FC in the DMN in cognitively impaired patients only compared to HCs,

one study [124] indicated an increase in FC in DMN in CIS compared to RRMS, one study [135] showed increased FC in no-pain patients versus patients who experienced pain and one study [13] observed an increase in DMN FC after a computer-based cognitive rehabilitation program. Motor networks were found to be affected in four studies [35, 41, 68, 124]. One study [35] only showed a correlation between EDSS score and FC between motor resting-state network and dorsal pre-motor cortex (PMd). One study [41] observed an increase in motor network FC which also correlated with the activation level of an area activated during a separate task. Another study [124] demonstrated an increase in CIS participants only and one study [68] found a decrease in FC in MS compared to HCs, which was also negatively correlated with disease severity. The frontoparietal network (FPN) was shown to be affected in four studies [28, 43, 114, 124]. One study [28] found decreased FC in the FPN in MS compared to HCs, another study [43] indicated increased FC, one study [114] showed an increase in FC in the FPN in MS participants who had undertaken a repetitive thumb flexions (RTF) exercise and another study [124] demonstrated an increase in FC in CIS patients compared to RRMS. Visual resting-state networks were affected in four studies [43, 50, 68, 114]. Two studies [43, 114] showed an increase in FC in visual networks in MS compared to HC and two studies [50, 68] found a decrease. One study [50] demonstrated increased FC between the visual resting-state network and extrastriate cortex in patients with previous optic neuritis and another study [68] showed disease severity to be positively correlated with medial visual network FC. The sensorimotor network (SMN) was affected in three studies [29, 43, 124]. All three studies found increased FC in this network in MS compared to HCs, though one of them [29] observed an increase in non-fatigued patients only and another study [124] indicated an increase in CIS patients only.

The executive function network was affected in two studies [114, 124]. One [114] found increased FC in this network after RTF exercises. Another study [124] showed FC in this network increased in MS compared to HC in CIS patients only. Auditory networks were found to be affected in one study [114] which demonstrated increased FC in MS compared to HC, and also showed FC to be increased post RTF compared to before RTF. The prefrontal insular resting-state network was observed to be increased in MS compared to HC in one study [43] which also found MS functional composite measure (MSFC) scores to negatively correlate with the FC of this network. The attention network had increased FC in MS compared to HC in one study [124], although this was for CIS only. CIS patients also expressed increased FC in this network compared to RRMS patients. The attention network was affected in two studies [114, 124]. One [124] found increased FC in MS compared to HC, though this was for CIS only. The same study also observed increased FC in this network in CIS compared to RRMS patients. The other study [114] demonstrated that this network only had increased FC in post RTF compared to before RTF. The cerebellar network (CBN) had increased FC in RRMS compared to HC in one study [114] which also showed an increase in FC in this network in post RTF compared to before RTF.

The control network was found to be affected in one study [68] which demonstrated a negative correlation between FC in this network and disease severity.

Seed to voxel

Six studies used a seed to voxel analysis [29, 31, 75, 131, 146, 157]; see Table 3.3 for results. The thalamus displayed significantly altered FC with different voxel clusters of the brain in MS compared to HC in four studies [31, 75, 131, 146]. One study [31] found only clusters with decreased FC located in the cerebellum, frontal and occipital cortices and the caudate nucleus. Another study [131] observed only increased clusters located in the sensorimotor cortex and ventral stream. Two studies [75, 146] showed clusters with both increased and decreased FC to the thalamus. Clusters with increased FC were located in the thalamus for one of the two studies [75] and in the cerebellum, basal ganglia, hippocampus, cingulum, bilateral temporal-occipital cortices, bilateral insular cortices, bilateral dorsal-frontal cortices and parietal cortex in the other study [146]. Clusters with decreased FC were found in the right middle frontal gyrus, right parahippocampal gyrus and left inferior parietal lobule for one study [75] and in the thalamus, cerebellum, cingulum, bilateral prefrontal cortices and bilateral parietal-occipital cortices in the other study [146]. A study [31] that compared participants who undertook a rehabilitation program (intervention group) to those who did not (control group) found clusters with both increased and decreased FC to the thalamus in the intervention group compared to the wait group. Disease duration was found to be negatively correlated with between-thalami FC in another study [75]. Expanded disability status scale (EDSS) score was negatively correlated with thalami eigenvector centrality mapping (ECM) values in one study [131]. In the other study [146] PASAT scores were negatively correlated with thalami FC.

The primary motor cortex (PMC) had significantly decreased FC with the left primary somatosensory cortex in MS compared to HC in one study [29]. Fatigue severity status scores were also negatively correlated with the FC between the PMC and supplementary motor area (SMA). One study [131] found clusters with both increased and decreased FC with a sensorimotor seed, increased clusters were located in the thalamus and decreased clusters in the ventral stream. The same study [131] showed both increased and decreased FC also with a ventral stream seed. Increased FC was observed in the thalamus and decreased FC in the sensorimotor cortex. The ECM values of a central stream seed were correlated with cognition in one study [146]. A posterior cingulate cortex (PCC) seed in MS compared to HC had decreased FC with other voxel clusters located in the anterior cingulate cortex (ACC) and right inferior frontal gyrus [157]. FC between a ventral medial pre frontal cortex (vmPFC) seed and a left frontal pole region was correlated with more stable performance on a semantic search reaction time test [157].

ROI analysis

Four studies used a region of interest (ROI) based analysis [35, 65, 132, 170]; see Table 3.4 for results. Expanded disability status scale was correlated with the connectivity between a dorsal premotor cortex (PMd) ROI and bilateral motor resting-state network signals in a study [35]. This correlation was only significant in RRMS, not SPMS. In another study [65], hippocampal ROIs (one left and one right) were correlated with AAL atlas time series ROIs, and they found increased FC in both hippocampi with the AAL in MS compared to HC. Memory performance was also correlated with increased FC of the left hippocampus. In another study [132], thalami ROIs were correlated with the AAL atlas time series where they found increased FC in MS with the thalami compared to inferior frontal gyrus, middle cingulate gyrus and inferior temporal lobe. Thalamic weighted degree was also increased in cognitively impaired versus cognitively preserved MS participants. Another study [170] used ICA to define DMN ROIs. These were based in the medial prefrontal cortex (MPFC), the left and the right inferior parietal lobules (IPL), the left and the right medial temporal lobes (mTL) and the posterior cingulate/precuneus (PCC/PCUN). All these ROIs showed increased FC, and none of them showed decreased FC. Paced auditory serial addition test (PASAT) scores were negatively correlated with the FC between the right mTL and PCC/PCUN.

AAL Atlas network measures

Two studies [65, 132] used the AAL as ROIs, which have been interpreted as ROI analysis rather than AAL analysis; see ROI analysis (Section 3.3.5). Of the studies that focused on the AAL all but one used the atlas to calculate different network measures. The one study that did not instead used it as part of a tool to classify fMRI data into either MS or HC [122]. Network measures found to be increased in MS compared to HC were nodal efficiency [42], local efficiency [42], connectivity index [42] and modularity [51]. Network efficiency and connectivity of 26 regions (in male MS vs male HC only) was found to be decreased in one study [133]. See Table 3.5 for results.

ALFF

ALFF was used in two studies [76, 171]. The bilateral thalami showed increased ALFF in both these studies with one of the studies [171] observing a correlation between PASAT and thalami ALFF and between mean fractional anisotropy and left thalami ALFF. Two other regions found to have increased ALFF in MS compared to HC were the right insula and the right superior temporal gyrus [76]. EDSS correlated with the ALFF of both these regions. See Table 3.6 for results.

Voxel to voxel

Voxel to voxel analysis was used in two studies. The first study [60] found increased FC in the DMN and control network. This study also found that cognitively efficient MS participants had greater FC in the control network and reduced FC in the DMN. The other study [172] found regions with both increased and decreased degree centrality and eigenvector centrality. See Table 3.7 for results.

3.3.6 Early MS

In three studies, neuroplasticity was found to be more prominent early on in disease progression. Roosendall *et al.* 2010 [124] found an increase of FC only in CIS which disappeared with disease progression. Liu *et al.* 2015 [75] found a negative correlation between disease duration and inter-thalamic FC. Gomboa *et al.* 2014 [51] and Faivre *et al.* 2012 [43] looked solely at early MS. All three studies suggest that functional reorganisation of the brain occurs predominately at an early stage of the disease. The reorganisation is a possible initial attempt by the brain to limit its functional atrophy.

3.3.7 Comparing MS to HC

Three studies did not have a comparison group, i.e. healthy controls; instead they compared MS patients before and after a cognitive task [13] or compared MS' FC with their clinical disability [35] or compared participants with chronic pain to those without [135]. Other studies whose main interest was not in comparing MS to HC still included HCs for extra comparison. For example, exploring how FC is altered by fatigue [29], how FC is altered by a cognitive rehabilitation program [31], how FC is altered by previous optic neuritis [50] or how FC in MS differs between genders [133]. The remaining studies all found differences between MS and HC, though their main aims were usually to explore how a clinical marker(s) of their choice is reflected through changes in FC.

3.3.8 Correlations with clinical markers

Conclusions about adaptive mechanisms in the brain were usually drawn by correlating a FC measure with a clinical marker. For example, Liu *et al.* 2015 [75] found a negative correlation between disease duration and between-thalami seed connectivity. They also found the between-thalami FC significantly greater in MS than HC. The authors concluded that their results suggests an adaptive role of the thalamus that is gradually lost as the disease progresses. Another example is the study by Dogonowski *et al.* 2013 [35] who demonstrate a positive correlation between EDSS score and FC between the motor resting-state network and the dorsal pre-motor cortex (PMd). This correlation was only present in RRMS, not SPMS, evidencing beneficial cortical adaptation with the integration of the PMd into the motor resting-state network.

3.4 Discussion

The most robust differences in FC between MS and HC groups were observed in the DMN, motor, visual and FPN networks as well as the thalamus. The DMN is one of the major networks associated with resting brain activity and is thought to represent internal cognition. The activation of the DMN is negatively associated with cognitive performance [135]. However, the findings from the studies included in the current review are inconsistent. Some studies show increased FC [41, 65], some show decreased FC [12] and some show both increases and decreases in FC [28] in the DMN for the MS groups compared to controls. Decreased FC may suggest disruption to the DMN, which is associated with cognitive impairment [12, 13, 28]. Increased FC, on the other hand, may suggest a form of compensation facilitated by neuroplasticity. However, the inconsistent findings across the literature make interpretations difficult. This inconsistency may be due to the heterogeneity in MS. MS is heterogeneous in terms of aetiology, pathology, symptomology and cognition. Further research using larger sample sizes are required to increase the reliability and validity of results. However, it may be possible that due to the heterogeneity there may be subtle differences in the neuroplasticity between MS populations, and that this process occurs on an individual level. I.e. the differences could cancel out when looking at a group level.

The findings of altered FC in visual and motor networks in MS is in line with the visual and motor disturbances associated with MS. This is further evidenced by the FC in these networks being correlated with disability. However, even in these networks the finding of FC is inconsistent, with some studies showing increases [41, 43, 114] and other studies decreases in FC [50, 68]. This inconsistency provides further evidence for the heterogeneity in MS, and it has some implications for the implementation of FC analysis. As even in well-defined networks, the results are inconsistent. This pattern of inconsistency has also been observed in the FPN network. This network is proposed as a key network for memory and cognition, where the parietal lobe integrates sensory information and then relays this to higher cognitive regions in the frontal lobe and is therefore particularly vulnerable to cognitive impairments [28]. Together the results do show neuroplasticity in MS. However the exact nature of the plasticity is inconsistent across the different studies and may adapt to the individuals' brain, thereby leading to inconsistent results in the literature.

The thalamus has multiple projections throughout the brain, through which it acts as a relay centre for a multitude of functions including motor [88, 137, 162], sensory [137, 162], and higher order cognitive functions such as memory [151] and attention [96, 113, 142, 143]. Furthermore, the thalamus is involved in other functions such as regulation of arousal in sleep and wakefulness [27, 129], where damage to the thalamus may result in a coma [10]. Given its widespread function, the thalamus' connectivity is important. However, the findings of FC in the

thalamus of the MS brains are inconsistent between studies. Some studies showed increased FC [75, 131, 146], whereas other studies showed decreased FC [31, 75, 146]. The inconsistency may be due to the large amount of heterogeneity in MS, where some participants are able to compensate for neuronal damage via neuroplasticity, whereas other participants are not. It may even be that the neuroplasticity adapts to the individuals' brain, leading to inconsistency across studies. Furthermore, due to its multiple connections it is difficult to determine whether the thalamus is directly involved in the pathological process, or whether it is susceptible to pathology from regions to which it is connected [92].

Even the most robust finding of FC changes in the DMN was only demonstrated in seven of the twenty-eight studies. This may be due to the fact that very few studies examine whole-brain connectivity, but rather choose specific regions on which to focus, based on previous studies. Each method has its advantages and disadvantages. Although a region-specific analysis may lead to more power despite smaller sample sizes, it may also miss key changes in the brain. Specifically in a group as heterogeneous as MS, where there is little consensus in the research, choosing regions based on previous studies seems ineffective. Whole-brain analyses can overcome this, however they often require much larger samples and therefore greater resources. The recent development of whole-brain topology measures [18] overcomes this limitation to some extent, by characterising local and global measures of a chosen network. However this again requires ROIs to be determined prior to the analysis.

Similarly, each analysis method has advantages and disadvantages and often the analysis technique employed relies upon the aim and hypotheses of a particular study. ICA is predominantly used as an exploratory tool, where there is no hypothesis to test. An advantage of ICA is that it is multivariate, needing no choice of region to explore; as a choice of region could be arbitrary or open to bias. ICA also requires no temporal model (i.e. the General Linear Model). Disadvantages are that it requires previous knowledge about functional brain networks to interpret the results. In fMRI, ICA is essentially on a voxel level, therefore care also has to be taken when looking for statistically different connectivity patterns between groups, for example multiple comparison and cluster size testing.

Seed and ROI based analysis have the advantage that their results are easier to interpret, because the selected seed/ROIs can be directly shown to be affected by MS. It also has a much lower dimensionality because the focus is on only a few seeds or at most a few hundred ROIs. This type of analysis is also spatially specific and able to provide information about how the connectivity between different regions of the brain compares between populations. However, it requires previous knowledge/expertise to select what regions to analyse, which is often difficult in a heterogeneous population as in MS. This is especially so for seed based analysis, which in

general focus on a lower (often singular) number of regions.

Network analysis using the AAL atlas has similar advantages and disadvantages to ROI based analysis, because the AAL atlas is essentially a segmentation of the brain into ROIs based on anatomy. The studies found in the review used the AAL almost exclusively for calculating network measures. These measures generally do not give spatial information but do highlight effects to the functional network topology of the brain. This is not a disadvantage to the AAL atlas itself, but to the type of analysis we found to be commonly deployed with it.

ALFF has the advantage of being a relatively simple metric to calculate and it gives signal information about a specific region. However, it does not elucidate the functional connectivity between different regions.

Finally, voxel to voxel correlation has the advantage of being multivariate, able to give information about connectivity between many regions and not requiring a priori information. However, it is computationally expensive with a large number of correlation values to calculate, and again care has to be taken when looking for statistically different connectivity patterns between groups.

Scan length has been shown to be an important variable for reliability of rs-fMRI results, with scan lengths of 9-13 minutes or longer greatly improving the technique [11]. The majority of studies found in this review do not meet this threshold. Therefore future studies should increase the total scanning time and volume number of their resting-state scans to increase the reliability of results. Furthermore, some studies claim to conduct rs-fMRI, however they only concatenate periods of rest between task-based fMRI [77].

Over 90% of individuals with MS experience fatigue [16]. Despite its subjective nature, it has a widespread impact on employment [48, 69], quality of life [67] and has significant cost implications [70]. However, only one study identified in this review focused on fatigue as their main aim, with the majority of studies focusing more on disability or cognitive function. Despite the widespread impact of fatigue there is a significant lack of research in this area. More research into the neural substrates of MS fatigue is required to improve our understanding and assessment of potential treatments to have real world benefit.

Most of the studies were cross-sectional and therefore do not take into account that MS participants have a degenerative disease and that their brains activity cannot be assumed to stay the same. More longitudinal studies could give helpful information as to how, and the rate at which, a MS patients' brain reorganises in response to the disease. Studies specifically following participants from onset of the disease could shed light on the effect of disease progression of MS

and expand the early MS results brought up in this review.

Due to the low number of studies, and large variety of focus of the studies, conclusions about their results are limited due to the lack of reproducibility. A benefit would be for future studies to try and replicate some previous findings, following the same methods but with a larger sample size of participants. This would give more validity and reliability to the results. A further, more practical, recommendation for future studies would be to include more detail of their preprocessing, stating more specifically what the steps were, rather than saying they just followed a standard protocol. The preprocessing protocol can have a major effect on the results of a study [136], affecting reliability and group discriminability. Inclusion of a detailed protocol, or even a template batch script, would greatly help in assessing the results from a given study.

3.5 Conclusion

Pure resting-state fMRI studies, observing neuroplasticity in MS were low in number and high in quality. The type of study, and methods used for analysis varied making a comparison of results difficult. This review identified six different methods used to assess neuroplasticity in the MS brains, each of which has specific advantages and disadvantages and were chosen based on the aims and hypotheses of the study. These various techniques display a range of different forms of evidence demonstrating alterations to the brain's functional organisation. The results showed evidence of neuroplasticity predominantly in the default mode network (DMN), motor network, visual network, frontal parietal network (FPN) and the thalamus. Although there are inconsistencies in whether this results in increased or decreased FC. The inconsistencies may be due to the heterogeneity in MS. It is recommended that studies with larger sample size are conducted to overcome the problem of heterogeneity and increase the reliability and reproducibility of the results.

Table 3.2: Summary of studies which used ICA to find evidence of neuroplasticity in MS versus HC.

Study	**Increased FC in MS vs HC**	**Decreased FC in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**
[12]	DMN	DMN	CP>CI: DMN CP<CI: DMN	240 volumes, TR=1508ms, voxel size=4x4x4mm
[13]	N/A	N/A	Increase after cCR: Between DMN and: PCC, Bilateral IPC Increase after aCT: No difference	240 volumes, TR=1508ms, voxel size=4x4x4mm
[29]	SMN (NF only)	None found	NF>F: SMN	270 volumes, TR=2000ms, voxel size= 3.5x3.5x4.02mm
[28]	None found	DMN (CI), Left FPN(CP,CI)	CP>CI: DMN, salience network. Global cognitive z-score correlated with Right-FPN (CP), salience (CP) FC	270 volumes, TR=2000ms, voxel size= 3.5x3.5x4.02mm

[35]	N/A	N/A	EDSS correlated with FC between bilateral motor resting-state network and PMd. Correlation still significant in RRMS, not in SPMS	480 volumes, TR=2490ms, voxel size= 3x3x3mm
[43]	Mean connectivity of non-dominant motor network	None found	Mean connectivity of non-dominant motor network significantly correlated with activation of area (right VLPC) recruited during separate task paradigm	250 volumes, TR=3600ms, voxel size= 2x2x2.5mm

[41]	Visual processing network, DMN, dorsal frontoparietal, prefronto-insular, right ventral frontoparietal, right sensorimotor	None found	Semantic fluency negatively correlated with posterior DMN FC. PASAT negatively correlated with dorsal frontoparietal and ventral frontoparietal FC. Left nine-hole peg test negatively correlated with left dorsal frontoparietal network FC. MSFC score negatively correlated with dorsal frontoparietal, right ventral frontoparietal and prefrontoinsular network FC	250 volumes, TR=3600ms, voxel size= 2x2x2.5mm
[50]	None found	Visual-RSN	Previous optic neuritis: Greater FC between visual-RSN and extrastriate cortex	240 volumes, TR=1508ms, voxel size= 4x4x4mm

[68]	None found	Motor network, medial visual network, lateral visual network	Disease severity negatively correlated with motor network and executive control network FC. Disease severity positively correlated with medial visual network FC	180 volumes, TR=2000ms, voxel size=3.44x3.44x 3.44mm
[114]	Auditory RSN, Visual RSN	None found	Post RTF >Pre RFT: SMN, CBN, auditory, right and left frontoparietal, ventral-dorsal attention, executive function network Increase in CBN significantly greater in RRMS vs HC	140 volumes, TR=3000ms, voxel size=not specified
[124]	Executive function network (CIS only), attention system (CIS only), sensorimotor function network (CIS only), left premotor cortex (CIS only), supplementary motor area (CIS only)	None found	CIS >RRMS: Attention system, DMN, left and right FPN	200 volumes, TR=2850ms, voxel size=3.3x3.3x3.3mm

[135]	N/A	N/A	No-pain >pain: DMN	100 volumes, TR=3000ms, voxel size=3x3x3mm
[170]	DMN (ICA used to define ROIs)	None found	PASAT negatively correlated with DMN ROIs FC	240 volumes, TR=2000ms, voxel size=3.125x3.125x 4mm

FC: Functional connectivity, DMN: Default mode network, CBN: cerebellar network, CP: Cognitively preserved, CI: Cognitively impaired, cCR: Computer-based cognitive rehabilitation, aCT: Aspecific cognitive rehabilitation, PCC: Posterior cingulate cortex, IPC: Inferior cingulate cortex, SMN: Sensorimotor cortex, F: Fatigued, NF: Not Fatigued, FPN: Fronto parietal network, EDSS: Expanded disability status scale, PMd: Dorsal pre-motor cortex, RRMS: Relasping Remitting MS, SPMS: Secondary Progressive MS, VLPC: Ventrolateral prefrontal cortex, PASAT: Paced auditory serial addition test, MSFC: Multiple sclerosis functional composite measure, RSN: Resting-state network, RTF: Repetitive thumb flexions, CIS: Clinically isolated syndrome

Table 3.3: Summary of studies which used a seed to voxel analysis to find evidence of neuroplasticity in MS versus HC.

Name	**Seed(s) with increased FC clusters in MS vs HC**	**Seed(s) with decreased FC clusters in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**
[28]	None found	**Left PMC seed:** left primary somatosensory cortex	FSS scores negatively correlated with FC between PMC and SMA	270 volumes, TR=2000ms, voxel size = 3.5x3.5x4.02mm

[31]	None found	**Thalamus seed:** cerebellum, frontal and occipital cortices, caudate nucleus, thalamus	**Thalamus seed (Intervention group>control group):** cingulum, precuneus, bilateral parietal cortex **Thalamus seed (control group >Intervention group):** cerebellum, left prefrontal cortex	250 volumes, TR=3600ms, voxel size=2x2x2.5mm
[157]	**Left thalamus seed:** right thalami **Right thalamus seed:** left thalami	**Left thalamus seed:** right middle frontal gyrus **Right thalamus seed:** right parahippocampal gyrus, left inferior parietal lobule	Disease duration negatively correlated with FC between thalami	229 volumes, TR=2000ms, voxel size= 1.875x1.875x5mm
[131]	**Thalamus seed:** sensorimotor cortex, ventral stream **Ventral stream seed:** thalamus **Sensorimotor seed:** thalamus	**Ventral stream seed:** sensorimotor cortex **Sensorimotor seed:** ventral stream	EDSS negatively correlated with thalamus ECM. Cognition correlated with central stream ECM	202 volumes, TR=2200ms, voxel size= 3.3x3.3x3mm

[146]	**Thalamus seed:** cerebellum, basal ganglia, hippocampus, cingulum, bilateral temporo-occipital cortices, bilateral insular cortices, bilateral dorsal-frontal cortices, right parietal cortex	**Thalamus seed:** thalamus, cerebellum, cingulum, bilateral insular, bilateral prefrontal, bilateral parieto-occipital	PASAT at 3 seconds negatively correlated with FC between **thalamus seed** and: thalamus, cerebellum, right hippocampus, cortical areas in frontal, temporal, parietal and occipital lobes bilaterally PASAT at 2 seconds negatively correlated with FC between **thalamus seed** and: all regions for 3 seconds plus cingulum and left hippocampus	120 volumes, TR=3000ms, voxel size=not specified
[157]	None found	**PCC seed:** ACC, right inferior frontal gyrus	More stable performance on SSRT subtest correlated with greater FC between vmPFC seed and left frontal pole	150 volumes, TR=2000ms, voxel size=3.75x3.75x5mm

PMC: Primary motor cortex, FSS: Fatigue severity scale, SMA: Supplementary motor area, EDSS: Expanded disability status scale, PMd: Dorsal pre-motor cortex, PASAT: Paced auditory serial addition test, ACC: Anterior cingulate cortex, PCC: Posterior cingulate cortex, SSRT: Semantic search reaction time, vmPFC: Ventral medial prefrontal cortex

Table 3.4: Summary of studies which used a ROI analysis to find evidence of neuroplasticity in MS versus HC.

Name	**ROIs with increased FC in MS vs HC**	**ROIs with decreased FC in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**
[35]	N/A	N/A	EDSS correlated with FC between bilateral motor resting-state network and PMd. Correlation still significant in RRMS, not in SPMS	480 volumes, TR=2490ms, voxel size= 3x3x3mm
[65]	Left hippocampus, Right hippocampus	None found	Memory performance correlated with increased FC of left hippocampus	200 volumes, TR=2850ms, voxel size=3x3.3x3.3mm
[132]	Thalamus with: inferior frontal gyrus, middle cingulate gyrus, inferior temporal lobe	None found	CI>CP: thalamic weighted degree	202 volumes, TR=2200ms, voxel size= 3.3x3.3x3mm
[170]	MPFC, Left-IPL, Right-IPL, Left-mTL, Right-mTL, PCC/PCUN	None found	PASAT negatively correlated with FC between PCC/PCUN and right mTL	240 volumes, TR=2000ms, voxel size= 3.125x3.125x4mm

MPFC: Medial pre-frontal cortex, IPL: Inferior parietal lobules, mTL: Medial temporal lobe, PCC/PCUN: Posterior cingulate cortex/precuneus, PASAT: Paced auditory serial addition test, CI: Cognitively impaired, CP: Cognitively preserved

Table 3.5: Summary of studies which performed a network analysis with AAL ROIs to find evidence of neuroplasticity in MS versus HC.

Name	**Increased FC in MS vs HC**	**Decreased FC in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**
[42]	Nodal efficiency, local efficiency, connectivity index (mean degree)	None found	During follow up: Increase in EDSS negatively correlated with decrease in connectivity index. Decrease of MSFC values correlated with nodal efficiency. Increase of EDSS correlated with connectivity index. Increase in T2-LL correlated with mean local efficiency. Increase in T2-LL negatively correlated with mean nodal efficiency	208 volumes, TR=3310ms, voxel size=3.5x3.5x3.5mm
[51]	Modularity values	None found	Dual task performance negatively correlated with modularity	305 volumes, TR=2080ms, voxel size=3x3x2mm

[122]	N/A	N/A	Variation of random forest scheme used for classifying correlation matrices into either MS or HC. Correctly classified 18/22 MS and 12/14 HC. Misclassified MS had low legion load and low number of attacks	450 volumes, TR=1100ms, voxel size= 3.75x3.75x5.63mm
[133]	None found	Network efficiency, connectivity of 26 regions (male MS vs. male HC only)	Male HC>female HC: 51 regions. Female HC>male HC: Increased network efficiency. Male MS: Mean synchronisation likelihood correlated with visuospatial memory performance	200 volumes, TR=2850ms, voxel size= 3.3x3.3x3.3mm
[132]	Used AAL as ROIs with additional thalamus ROIs. See ROI table (Table 3.4).			
[65]	Used AAL as ROIs with additional hippocampal ROIs. See ROI table (Table 3.4).			

EDSS: Expanded disability status scale, MSFC: Multiple sclerosis functional composite, T2-LL: T2-Lesion load, CI: Cognitively impaired, CP: Cognitively preserved

Table 3.6: Summary of studies which used a ALFF analysis to find evidence of neuroplasticity in MS versus HC.

Name	**Increased FC in MS vs HC**	**Decreased FC in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**

[76]	Bilateral thalami, right insula, right superior temporal gyrus	None found	EDSS correlated with ALFF in right insular and right superior temporal gyrus region	170 volumes, TR=2000ms, voxel size= 1.875x1.875x5mm
[171]	Bilateral thalami	None found	PASAT correlated with left and right thalami ALFF. Mean fractional anisotropy correlated with left thalami ALFF	240 volumes, TR=2000ms, voxel size= 3.125x3.125x3mm

EDSS: Expanded disability status scale, ALFF: Amplitude of Low Frequency Fluctuations, PASAT: Paced auditory serial addition test

Table 3.7: Summary of studies which used a voxel to voxel analysis to find evidence of neuroplasticity in MS versus HC.

Name	**Increased FC in MS vs HC**	**Decreased FC in MS vs HC**	**Other comparisons with/within MS**	**Resting-state fMRI protocol**
[60]	DMN, Control network	None found	Greater FC in DMN in cognitively less-efficient participants. Greater FC in control network in cognitively efficient MS	606 volumes, TR=2000ms, voxel size=2x2x2mm

[172]	Regions with increased degree centrality and eigenvector centrality (See paper)	Regions with decreased degree centrality and eigenvector centrality (See paper)	EDSS negatively correlated with decreased degree centrality in bilateral M1/S1 (relapsing phase). TWMLL negatively correlated with decreased degree centrality in bilateral M1/S1 (relapsing phase). BPF correlated with decreased DC in left OP, left insular, SMA (remitting phase). EDSS correlated with increased DC in right M1 (remitting phase). EDSS correlated with decreased EC in left IPL (remitting phase)	240 volumes, TR=2000ms, voxel size= 3.125x3.125x4mm

DMN: Default mode network, EDSS: Expanded disability status scale, M1: Primary motor cortex, S1: Primary somatosensory cortex, TWMLL: Total white matter lesion load, BPF: Brain parenchymal fraction, DC: Degree centrality, EC: Eigenvector centrality, OP: Operculum parietal, SMA: Supplementary motor area, IPL: Inferior parietal lobule

Table 3.8: Summary of participant demographics in the studies.

Study	**Number of MS participants**	**Number of HC participants**	**Other details**
[12]	36 RRMS	18	Matched for sex, age and education
[13]	32 RRMS	0	RRMS split into two groups, one receiving cognitive rehabilitation, the other not

[29]	60 RRMS	18	RRMS classified as fatigued (n=32) and not fatigued (n=28)
[28]	60 RRMS	18	RRMS classified as cognitively impaired (n=30) and cognitively preserved (n=30). HC all right handed
[31]	24 RRMS	11	
[35]	27 RRMS	0	
[41]	13 early RRMS	14	Matched for sex and age
[43]	13 early RRMS	14	Matched for sex, age and education
[42]	38 RRMS	24	
[50]	30 RRMS	15	RRMS classified as with optic neuritis (n=14) and without optic neuritis (n=16)
[51]	8 RRMS, 8 CIS	20	7 RRMS and 8 CIS for session 3
[60]	12 RRMS, 2 CIS, 2 MS	16	Matched for sex, age and education
[65]	40 RRMS, 17 SPMS	28	Matched for sex, age and education
[68]	28 RRMS	28	Matched for sex, age and education
[75]	35 RRMS	35	Matched for sex and age
[76]	35 RRMS	35	Matched for sex and age
[114]	20 RRMS	14	Matched for sex and age. RRMS all right handed
[122]	22 RRMS	14	
[124]	31 RRMS, 14 CIS	41	Matched for sex and age
[131]	112 RRMS, 7 PPMS, 9 SPMS	50	Matched for sex, age and education
[132]	133 RRMS, 15 PPMS, 9 SPMS	47	

[133]	30 CDMS	30	Matched for age, education and pre morbid IQ. Male and Female MS (12 male RRMS, 3 male SPMS, 14 female RRMS, 1 female SPMS) matched for disease duration, disability and lesion load
[135]	14 RRMS, 8 SPMS, 1 PPMS	0	12 MS with chronic pain (7 RRMS, 4 SPMS, 1 PPMS) and 11 without (7 RRMS, 4 SPMS, 0 PPMS) matched for sex, age, disease duration and EDSS
[146]	55 RRMS	24	
[157]	18 RRMS	16	All female, matched for age and education
[170]	24 RRMS	24	Matched for sex, age and education
[171]	23 RRMS	23	Matched for sex and age
[172]	34 RRMS	34	Matched for sex, age and education

RRMS: Relapsing-remitting MS, PPMS: Primary-progressive MS, SPMS: Secondary-progressive MS, CDMS: Clinically definite MS, CIS: Clinically isolated syndrome, EDSS: Expanded disability status scale

Table 3.9: Quality Assessment.

Study	Research Objective	Recruitment Procedure	Inclusion/ exclusion	Population demo-graphics	Imaging protocol	Comparison group	Pre process-ing protocol
[12]	Y	Y	Y	Y	Y	Y	N
[13]	Y	Y	Y	Y	Y	N	N
[29]	Y	Y	Y	Y	Y	Y	Y
[28]	Y	Y	Y	Y	Y	Y	Y
[31]	Y	Y	Y	Y	Y	Y	Y
[35]	Y	Y	Y	Y	Y	N	Y
[41]	Y	Y	Y	Y	Y	Y	Y
[43]	Y	Y	Y	Y	Y	Y	Y
[42]	Y	Y	Y	Y	Y	Y	Y
[50]	Y	Y	Y	Y	Y	Y	Y(W)
[51]	Y	Y	Y	Y	Y	Y	Y
[60]	Y(W)	Y	Y	Y	Y	Y	Y
[65]	Y	Y	Y	Y	Y	Y	Y
[68]	Y	Y	Y	Y	Y	Y	Y
[75]	Y	Y	Y(W)	Y	Y	Y	Y
[76]	Y	Y	Y(W)	Y	Y	Y	Y
[114]	Y	Y	Y	Y	Y	Y	Y
[122]	Y	Y	Y	Y	Y	Y	Y
[124]	Y	Y	Y(W)	Y	Y	Y	Y
[131]	Y	Y	Y	Y	Y	Y	Y
[132]	Y	Y	Y(W)	Y	Y	Y	Y
[133]	Y	Y	Y	Y	Y	Y	Y(W)
[135]	Y	Y	Y	Y	Y(W)	N	Y
[146]	Y	Y	Y	Y	Y	Y	Y
[157]	Y	Y	Y	Y	Y	Y	Y
[170]	Y	Y	Y	Y	Y	Y	Y
[171]	Y	Y	Y	Y	Y	Y	Y
[172]	Y	Y	Y	Y	Y	Y	Y

Y: Criteria included, N: Criteria not included, Y(W): Criteria included though weakly explained

CHAPTER 4

Analysing Large Scale Functional Networks in Multiple Sclerosis

The brain is widely regarded as one of the most complex objects known, with many interacting components. Due to this complexity, our understanding of how the brain works and how it can be affected is limited. As the brain is composed of many interconnected parts, viewing the brain as a network is a natural way in which it can be analysed to further our understanding [46].

Many disorders affect the brain, one of particular interest is Multiple Sclerosis (MS), a demyelinating disease of the central nervous system. MS is a leading cause of disability in young adults and its prevalence around the world is increasing [17]. Using functional magnetic resonance imaging (fMRI), an indirect measure of activity within the brain, we can create functional networks and see if properties of these networks differ between MS and Healthy control (HC) participants. This could elucidate how the disease progresses and inform on the effectiveness of treatments.

fMRI is a method which uses MRI to indirectly measure activity within the brain. Oxygenated blood behaves differently under a magnetic field than de-oxygenated blood, and an MRI scanner can use this to measure blood flow around the brain, recording a blood oxygen level dependent (BOLD) signal. As brain cells require oxygen to function, and when more active require more oxygen, the BOLD signal is an indirect measure of activity within the brain.

Traditionally, fMRI has been used to see which areas of the brain are more active during a task. More recently there has been a growing interest in resting-state fMRI [119], a way of

measuring the intrinsic connectivity between parts of the brain, termed functional connectivity (FC), in its basal state. Alterations in FC between MS and HC participants could be evidence of neuroplasticity in the MS brains, functioning as a compensatory mechanism. The review of Chapter 3 describes studies we found which looked at evidence of neuroplasticity in MS from resting-state fMRI studies. The main conclusions from this review are that MS is highly heterogeneous, reflected in the conflicting results reported, and that there is a lack of research into one of the main effects from the disease, fatigue.

There is a prevalence of fatigue in MS [16], and this is the main reason the study of this chapter was carried out. The analysis we performed can be broadly split into two parts; hypothesis driven and exploratory. The hypothesis is there is disruption to the basal ganglia caused by MS; the basal ganglia is a subcortical structure of the brain whose disruption has been assiciated with fatigue [22]. One of the symptoms of this disruption may be the fatigue experienced by the MS participants. The exploratory analysis we carried out was looking at different global network measures of the FC between regions of the brain. We looked at a number of different measures with no prior expectation of exactly how they might differ between the two populations. Disruption to a network measure could be evidence of compensatory neuroplasticity, which could indirectly explain the MS participants fatigue, as a different FC structure may use more metabolic energy.

4.1 Background of study

Participants were recruited by CRICBristol, who recruited 40 HC and 38 MS participants. They all performed a number of tests whilst inside the scanner, i.e. task-based fMRI, but also spent a period of the scanning at rest, i.e. resting-state fMRI. The analysis in this chapter is looking solely at the resting-state fMRI. Resting-state designs require no engagement from participants, require no experimental design and are an increasingly popular way of investigating basal brain activity. The resting-state scan we used had a time to repetition (TR) of 906ms (i.e. a sampling frequency of $\frac{1}{0.906}$) and recorded 300 volumes, with a total resting-state scanning time of 4.53 minutes. This is shorter than the recommended 9-13 minutes or longer as highlighted in the review of Chapter 3. The main reason for this was the study paradigm performed by CRICBristol involved tasks and a period of resting-state, and time constraints in the scanner resulted in a compromise between the two.

The reason CRICBristol performed this study was because of the aforementioned prevalence of fatigue in MS, and differences found between HC and MS could explain what may be causing this fatigue. We used a network analysis, abstracting regions of the brain as nodes of a network, and the correlation between the regions activity as the edges of the network. We measured

different properties of these networks and compared them between the two populations to test for a difference in the functional workings of the MS populations brains.

4.2 Study participants

CRICBristol initially recruited 40 female participants (aged 32-67, mean 49.8) with a clinical diagnosis of MS, from The Brain Centre at Southmead Hospital, North Bristol Foundation Trust, UK. Diagnosis were made according to the McDonald criteria [89, 115, 116]. The 40 female HC participants (aged 31-68, mean 49.5) were matched for age (and gender). A technical problem on the computer system used to store the study data resulted in two of the MS participants data being corrupted; resulting in the final 38 MS participants (aged 32-67, mean 49.9).

Of the 38 MS participants, 25 were diagnosed with relapsing-remitting MS (RRMS), 4 with primary progressive MS (PPMS) and 9 with secondary progressive MS (SPMS) See Table 4.1 for MS participant demographics (diagnosis, age and disease duration).

Diagnosis	**Number of participants**	**Age (Years)**	**Disease duration (Years)**
Relapsing-remitting MS (RRMS)	25	46.8 (32-67)	8.96 (1-29)
Primary progressive MS (PPMS)	4	57 (52-59)	9.25 (6-14)
Secondary progressive MS (SPMS)	9	55.2 (40-65)	23.6 (12-36)

Table 4.1: Table detailing breakdown of MS particpants by diagnosis, age and disease duration. Mean (Range).

4.3 Preprocessing

Before any comparison between groups can be made, the images acquired need to be cleaned up and prepared for analysis in a process called preprocessing. This is because fMRI is inherently noisy and participants do not keep perfectly still during the scanning, among other reasons discussed in Sections 4.3.1 and 4.3.2. Preprocessing was performed in SPM12 [112] (see below) within MATLAB and further preprocessing was performed using custom code in MATLAB.

4.3.1 SPM preprocessing

Statistical parametric mapping (SPM) is a MATLAB toolbox developed by members and collaborators of the Wellcome Centre for Human Neuroimaging. SPM12 was the version we used for

preprocessing and analysis of the fMRI data, though it is also capable of analysing data from PET, SPECT, EEG and MEG as well. We only used the preprocessing parts of SPM12, performing further preprocessing and analysis separately in MATLAB. During the scanning a structural image is acquired in addition to the BOLD functional scans, this is used by SPM as part of the preprocessing.

The first step is realignment, which uses a least squares approach to a rigid body 6-parameter spatial transformation to calculate the head movement of the participant. Figure 4.1 shows the calculated movement parameters of a HC participant, Figure 4.1a showing the x, y and z translation motions, and Figure 4.1b showing the pitch, roll and yaw rotation motions. Scans are then adjusted, using these calculated movement parameters, to realign them with the first image in the series.

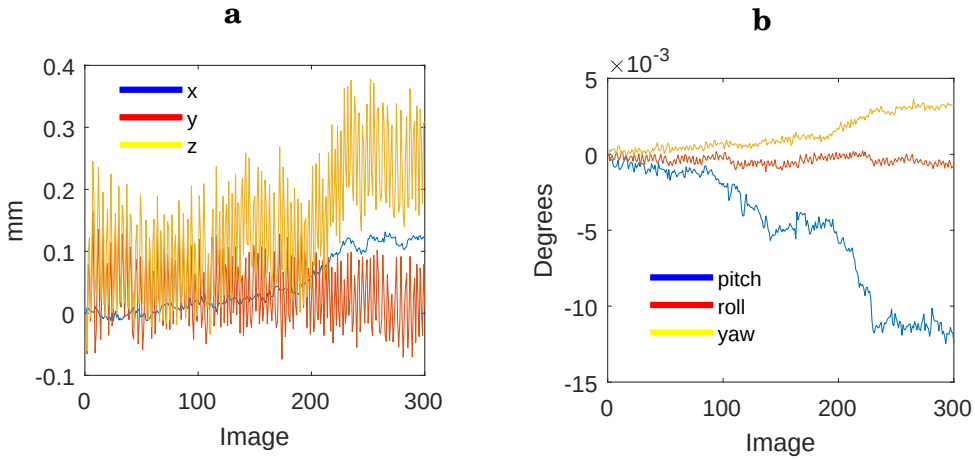


Figure 4.1: Calculated movement parameters of a single HC participant. **a** Translation movement. **b** Rotation movement.

The next step is corregistration. This procedure uses the structural scan as a reference image fixing it in place, whilst jiggling the mean of the functional scans such that the mutual information between the two is maximized. The rest of the functional scans are altered the same as their mean, bringing all the functional images in alignment with the structural image. After this, segmentation is performed on the structural image. An original structural image is shown in Figure 4.2a, with grey matter (GM) segmented shown in Figure 4.2b, white matter (WM) in Figure 4.2c and cerebral spinal fluid (CSF) in Figure 4.2d.

Everyone's brain is different, normalization is the process by which the images are deformed such that they all fit a standard template. Fitting everyone's brain to a template ensures comparable anatomical structures in the brain match up between participants.

The template chosen is the Montreal Neurological Institute (MNI) space. All the BOLD images were normalized, along with the segmented GM, WM and CSF images. Bounding box

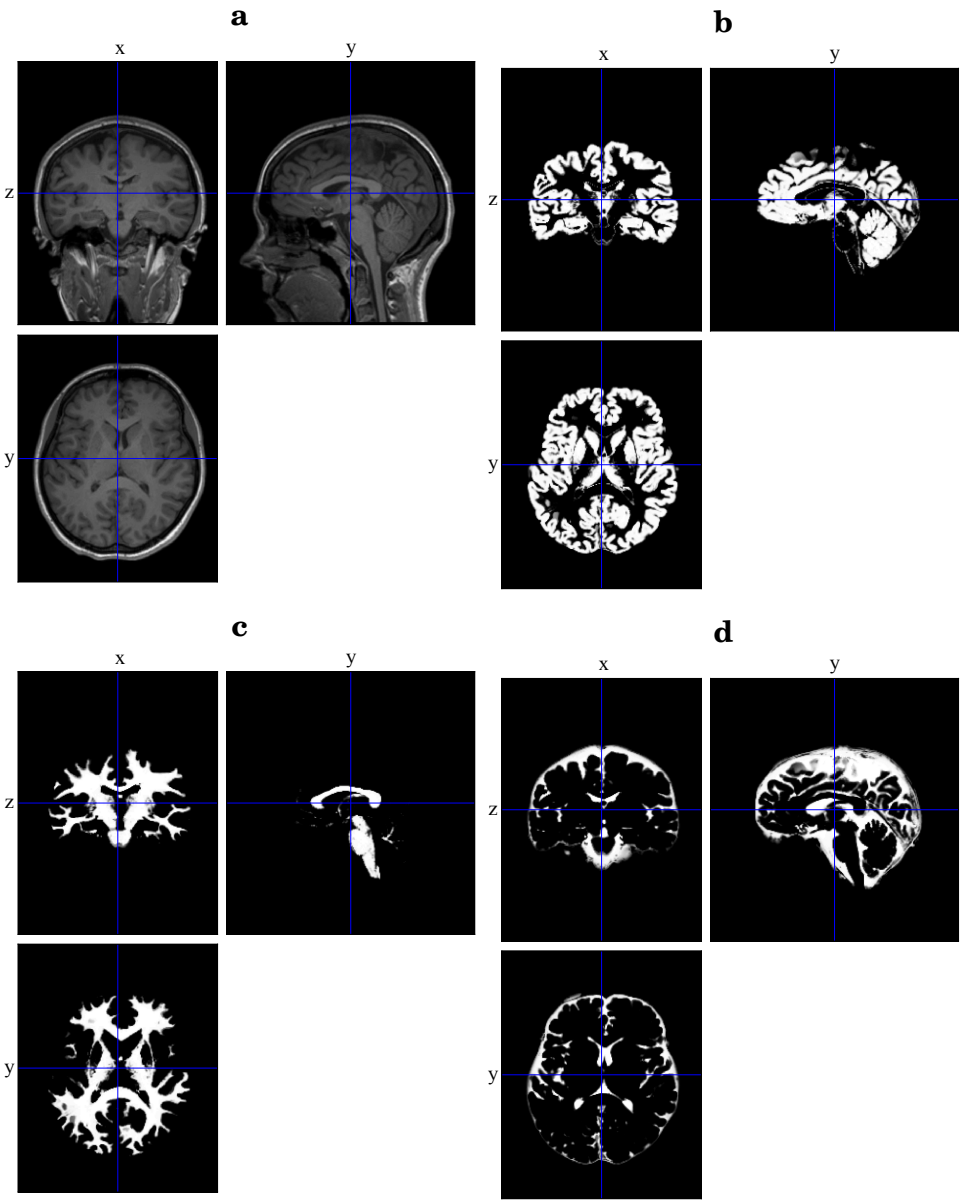


Figure 4.2: SPM segmentation of the HC participant. **a** Structural image. **b** Segmented grey matter image. **c** Segmented white matter image. **d** Segmented cerebral spinal fluid image.

settings were changed from defaults to match the AAL template (See Section 4.4 for AAL details).

The final process we used SPM for before exporting into 4D matrices for use within Matlab is smoothing. The previous steps above are not perfect, and so smoothing is used to increase the overlap areas of the brain and ensure a better match of anatomical regions between participants. This uses a 3D Gaussian kernel with a full width at half maximum size of 8x8x8 mm^3. This is performed on all the normalized functional images. Figure 4.4 shows a smoothed BOLD image (compare to the unsmoothed normalized image in Figure 4.3b).

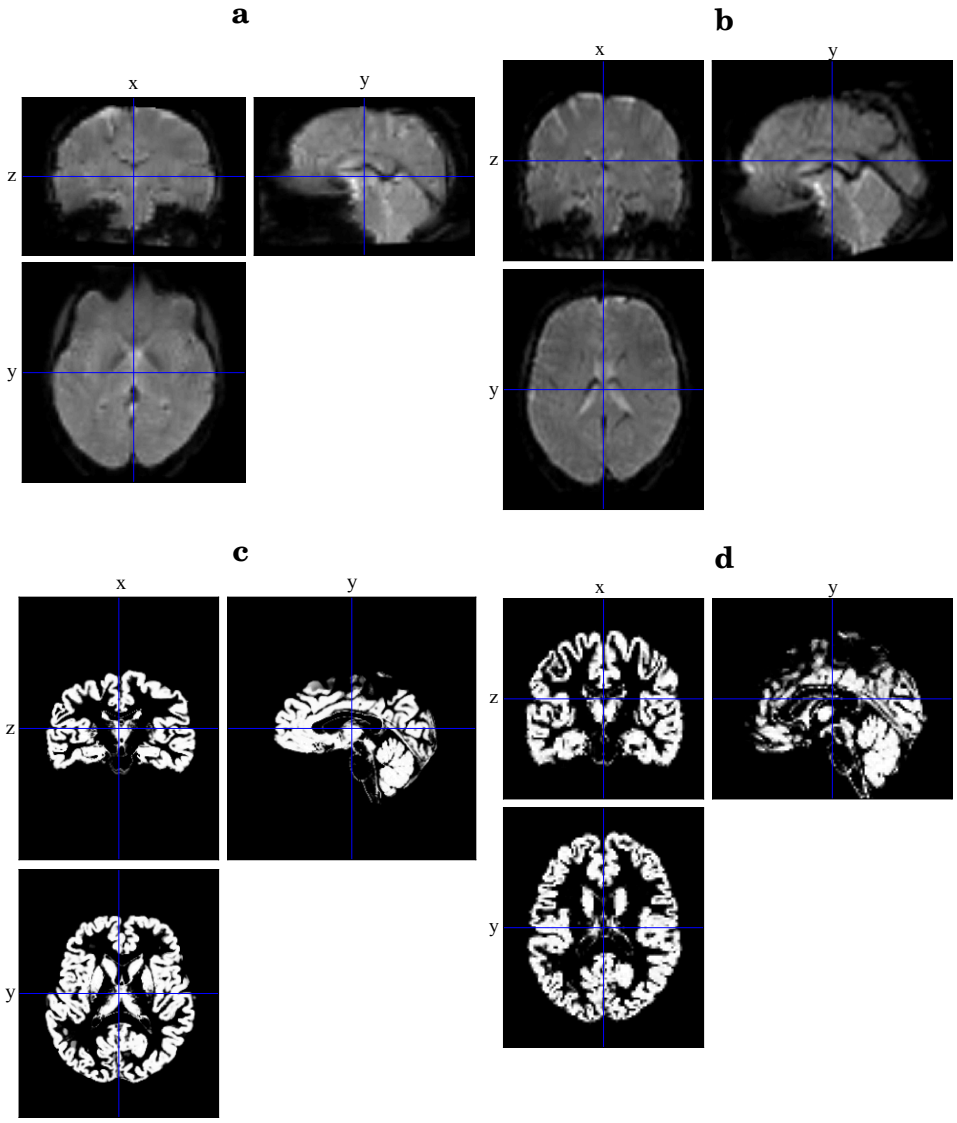


Figure 4.3: SPM normalization of the HC participant. **a** Realigned BOLD image. **b** Normalised BOLD image. **c** Segmented GM image. **d** Normalized segmented GM image.

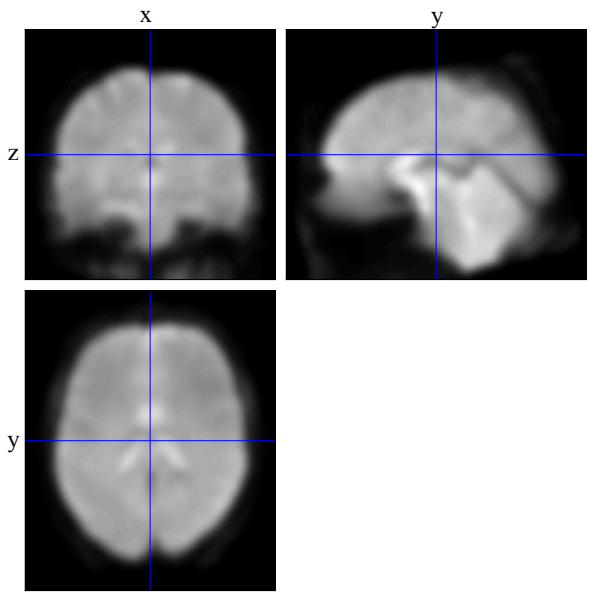


Figure 4.4: Smoothed BOLD image of the HC participant.

4.3.2 Further preprocessing in MATLAB

We converted all the smoothed functional images into a single 4D matrix for each participant. Time series for predefined regions of interest (ROI) were extracted by averaging the BOLD signals from all the voxels that fall within the ROI. See Section 4.4 for details on selection of ROIs. Certain signals affect the measurement of the BOLD signal and need to be accounted for. These signals are the movement of the participant, the BOLD signal from white matter regions and the BOLD signal from CSF regions. The movement signals used were obtained from SPMs realignment step (See Figure 4.1 for an example of movement signals). White matter signal was obtained from averaging all the BOLD signals from the voxels that fall within the segmented white matter image (see Figure 4.2c for an example of a (pre-normalized) white matter mask) and similarly for the CSF signal (Figure 4.2d shows an example of a (pre-normalized) CSF mask). These nuisance signals were removed using linear regression, see Section 4.3.3 for details of this.

The final preprocessing step is bandpass filtering. In the Systematic Review of Chapter 3, we found 22/28 of the studies used a bandpass filter, with half of these using only a highpass filter (11/22 of the studies). This removes low frequency noise attributed to properties of the MRI scanner (termed scanner drift). However it does not remove high frequency noise attributed to physiological signals like breathing and blood pulse, and it has been demonstrated that connectivity is predominantly characterized by frequencies lower than these [25]. We therefore chose the most popular bandpass range which included a highpass and a lowpass cut-off of 0.01-0.08 Hz (6/22 of the studies). The remaining studies from the review (5/22) each chose different bandpass filter values, with varying highpass and lowpass thresholds. Figure 4.5 shows two examples of extracted BOLD time series for two different ROIs, and those respective time series after bandpass filtering.

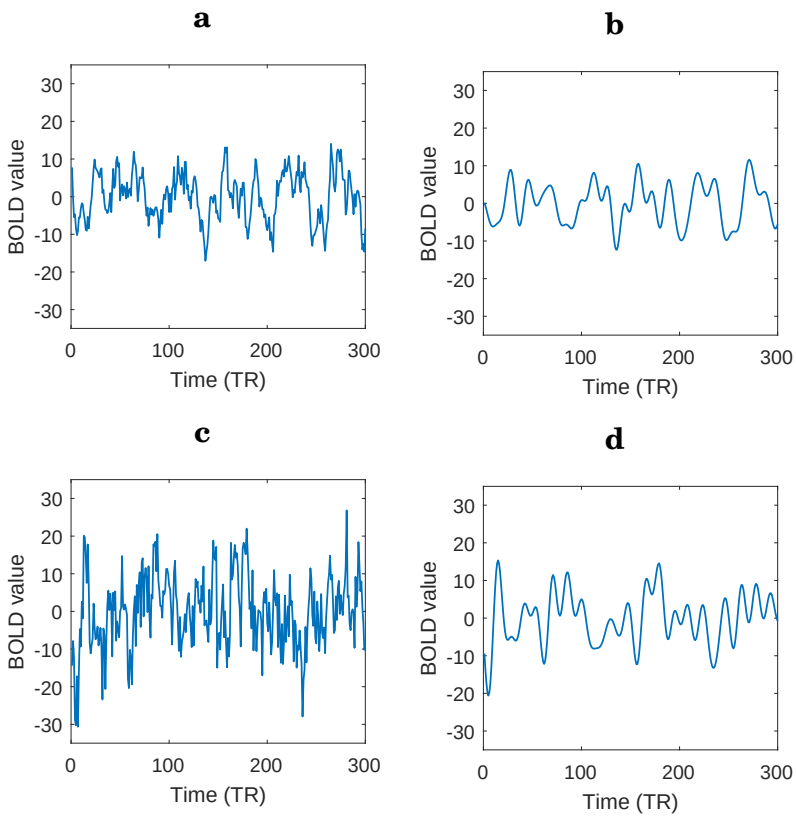


Figure 4.5: Example of bandpass filter applied to extracted time series from AAL ROIs for the single HC participant demonstrated in Section 4.3.1. **a** Extracted time series of AAL region 1 (left precentral gyrus). **c** AAL region 71 (left Caudate). **b**, **d** Series bandpassed filtered respectively, with filter range of 0.01-0.08 Hz.

4.3.3 Equivalence of linear regression and time series extraction ordering

The standard way of nuisance regression in the brain imaging community is to perform nuisance regression on each voxel of the brain individually and then extract time series of averages of the voxels that fall within a ROI [100]. As there are approximately $O(10^6)$ voxels in a fMRI brain image this is a potentially computationally heavy procedure. If later on in the analysis it is decided to use different or modified nuisance signals, performing the regression multiple times could take a long time. Extracting the time series from voxels and then performing the regression on the extracted time series would greatly reduce computational load, as there is a maximum of 264, $O(10^2)$, ROIs. It would also be more flexible, as the time series are already extracted, ready for analysing.

The regression used is linear, which suggests that swapping the order of regression and extraction would have no effect. However it is not immediately clear this is the case. Therefore, we re-derived the linear nuisance regression in terms of how it is used on the fMRI data and

prove mathematically that it is equivalent to extract ROIs and then regress nuisance signals from those, rather than nuisance regression on every voxel first. The form of the linear regression is given by

$$\begin{bmatrix} b(t=1) \\ b(t=2) \\ \vdots \\ b(t=T) \end{bmatrix} = \begin{bmatrix} N_1(t=1) & N_2(t=1) & \cdots & N_n(t=1) \\ N_1(t=2) & N_2(t=2) & \cdots & N_n(t=2) \\ \vdots & \vdots & \vdots & \vdots \\ N_1(t=T) & N_2(t=T) & \cdots & N_n(t=T) \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + \begin{bmatrix} \epsilon(t=1) \\ \epsilon(t=2) \\ \vdots \\ \epsilon(t=T) \end{bmatrix} \tag{4.1}$$

where the column vector on the left represents the BOLD time signal for a single voxel or ROI after initial preprocessing, with the value $b(t)$ being its BOLD signal value measured at time t. The matrix of regressor signals is composed of column vectors $N_i, i \in [1,n]$, where each column vector is the time series of nuisance signal i, with there being n nuisance regressors in total. This regressor matrix multiplies by its weighting column vector, composed of the weightings $w_i, i \in [1,n]$ determined by least squares. The final column vector is the error column vector. This error vector is the signal we are interested in, as it is what is left of the signal after removing the nuisance signals. The linear regression equation in a more compact form is given by

$$\mathbf{b} = \mathbf{N}\mathbf{w} + \epsilon. \tag{4.2}$$

The aim of the linear regression is to minimise the error ϵ. The most common method, and the one used here, is by using least-squares. The derivation starts with rearranging Equation (4.2) to make the error the subject,

$$\epsilon = \mathbf{b} - \mathbf{N}\mathbf{w}. \tag{4.3}$$

We sum the square of the errors and simplify,

$$\mathbf{S} = \sum_i^T \epsilon^2 = \epsilon^\top \epsilon = (\mathbf{b} - \mathbf{N}\mathbf{w})^\top (\mathbf{b} - \mathbf{N}\mathbf{w}) = (\mathbf{b}^\top - \mathbf{w}^\top \mathbf{N}^\top)(\mathbf{b} - \mathbf{N}\mathbf{w}) ==$$

$$= \mathbf{b}^\top \mathbf{b} - \mathbf{w}^\top \mathbf{N}^\top \mathbf{b} - \mathbf{b}^\top \mathbf{N}\mathbf{w} + \mathbf{w}^\top \mathbf{N}^\top \mathbf{N}\mathbf{w}$$

$$= \mathbf{b}^\top \mathbf{b} - 2\mathbf{b}^\top \mathbf{N}\mathbf{w} + \mathbf{w}^\top \mathbf{N}^\top \mathbf{N}\mathbf{w}. \tag{4.4}$$

By differentiating equation 4.4 with respect to $\mathbf{w}$, we obtain

$$\frac{\partial \mathbf{S}}{\partial \mathbf{w}} = -2\mathbf{N}^\top \mathbf{b} + 2\mathbf{N}^\top \mathbf{N}\mathbf{w}, \tag{4.5}$$

which we set to 0 to find the minimum, i.e.

$$\min \mathbf{S} \quad \text{is where} \quad \frac{\partial \mathbf{S}}{\partial \mathbf{w}} = 0. \tag{4.6}$$

This results in the following expression:

$$\mathbf{N}^\top \mathbf{N}\mathbf{w} = \mathbf{N}^\top \mathbf{b}. \tag{4.7}$$

A visual representation of this can be found by abstracting the multidimensional column space $\mathbf{N}$ onto a plane. Minimising the error in this case means finding $\mathbf{w}$ such that the error, ϵ, is perpendicular to the $\mathbf{N}$ column space plane, i.e. $\mathbf{N}^\top \epsilon = 0$. This can thought of as taking multiple simultaneous vector dot products of ϵ with all the components of $\mathbf{N}$ and is illustrated in Figure 4.6.

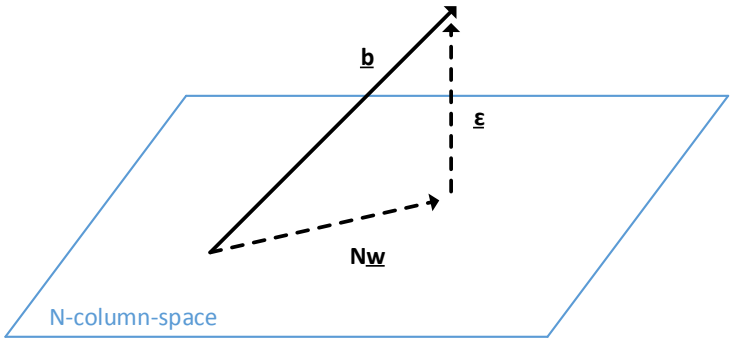


Figure 4.6: Column space representation of least squares regression. With BOLD signal vector $\mathbf{b}$, nuisance signals $\mathbf{N}$, weighting vector $\mathbf{w}$ and error vector ϵ.

Using the column-space representation (Figure 4.6) we obtain an expression to solve,

$$\mathbf{N}^\top \epsilon = 0. \tag{4.8}$$

Substituting the error from equation (4.3) results in

$$\mathbf{N}^\top (\mathbf{b} - \mathbf{N}\mathbf{w}) = 0. \tag{4.9}$$

Multiply the brackets

$$\mathbf{N}^\top \mathbf{b} - \mathbf{N}^\top \mathbf{N}\mathbf{w} = 0, \tag{4.10}$$

and simplify, resulting in an equivalent expression to finding the minimum of the error;

$$\mathbf{N}^\top \mathbf{N}\mathbf{w} = \mathbf{N}^\top \mathbf{b}. \tag{4.11}$$

Now to define the error vector in terms of the measured signal, **b**, and the Nuisance regressor matrix, **N**. Multiply both sides of equation (4.7) (and equivalently equation (4.11)) by the inverse of $\mathbf{N}^\top\mathbf{N}$ as shown by

$$(\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top\mathbf{N}\mathbf{w} = (\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top\mathbf{b} \tag{4.12}$$

which simplifies to

$$\mathbf{w} = (\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top\mathbf{b}. \tag{4.13}$$

By substituting this **w** from equation (4.13) into equation (4.3) we obtain

$$\epsilon = \mathbf{b} - \mathbf{N}(\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top\mathbf{b}, \tag{4.14}$$

which simplifies to

$$\epsilon = (\mathbf{I} - \mathbf{N}(\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top)\mathbf{b}. \tag{4.15}$$

We next define $\mathbf{K} = (\mathbf{I} - \mathbf{N}(\mathbf{N}^\top\mathbf{N})^{-1}\mathbf{N}^\top)$. Then, we obtain

$$\epsilon = \mathbf{K}\mathbf{b}, \tag{4.16}$$

where the matrix **K** is a constant, defined by the nuisance regressor matrix.

We want to show that regressing every voxel and then extracting signals from them is equivalent to first extracting signals and then regressing the extracted signals. We do this by considering voxels we want to extract and regress, and term them b_m where m ranges from 1 to M. The first case considered is regressing each voxel separately, and then extracting the average of them. Each voxel is regressed separately and is given by $\epsilon_m = \mathbf{K}b_m$. The average of the regressed voxels is given by

$$B_{ROI} = \frac{1}{M}\sum_1^M \epsilon_m = \frac{1}{M}\sum_1^M \mathbf{K}b_m = \frac{\mathbf{K}}{M}\sum_1^M b_m. \tag{4.17}$$

The second case is where the average of the voxels is taken and then regressed. The average of the voxels is given by $\frac{\mathbf{1}}{M}\sum_1^M b_m$, and regressing these given by

$$B_{ROI} = \mathbf{K}\cdot\frac{\mathbf{1}}{M}\sum_1^M b_m = \frac{\mathbf{K}}{M}\sum_1^M b_m. \tag{4.18}$$

Equations (4.17) and (4.18) are mathematically equivalent. Therefore it is more computationally efficient to first extract ROI time series, and then to perform nuisance regression. We apply nuisance regression to our analysis in the more efficient way described here.

4.4 Regions of interest (time series extraction)

We used a region of interest (ROI) analysis to explore neuroplasticity via functional connectivity (FC) with three different sets of ROIs considered, termed Fair, Cole and AAL. All three system of ROIs encompass the whole brain, enabling the exploratory analysis of large scale network disruption. The AAL contains ROIs belonging to the basal ganglia, enabling the testing of the disrupted basal ganglia hypothesis.

Fair and Cole base their ROIs inspired by a large collation of evidence which suggests the brain can be subdivided into functional networks, shown to be involved in distinct processes within the brain [90]. For Fair and Cole we define their ROIs as spheres located as stated in their studies, with radii of 4mm. We extracted time series as the mean of all the voxel BOLD time series that fall within each respective ROI sphere.

Fair based their networks on previous studies, manually collating positions for all the nodes from their networks of interest. They define 30 ROIs belonging to 3 distinct networks: the fronto-parietal network (FPN), the cingulo-opercular network (CON) and the default mode network (DMN). The FPN and CON are involved in task control [37, 38], where the FPN supports top-down (in the moment) maintenance of task performance whilst the CON primes and sustains the brain in a task state [36]. The DMN is comprised of regions consistently shown to be correlated at rest in adults, and is reported to be involved in autobiographical, self-monitoring value judgements and other self-referential mental activity [56, 90, 119].

Cole [24] is similar to Fair, i.e. based on resting-state functional networks, except it defines location of its nodes differently, and has more of them. Cole based the locations of their ROIs on a previous study [118] which performed an analysis on a large fMRI dataset to identify voxels that significantly changed when certain behaviours or stimuli were presented. The study combined this with a mapping process which treated the cortex of the brain as a 2D sheet, and identifies areas where patterns in the BOLD signal rapidly change. These areas represent boundaries between different functional networks. Cole define 264 ROIs belonging to 11 distinct functional networks: the auditory network, CON, dorsal attention network (DAN), DMN, FPN, the salience network (SAN), the somatosensory network (SMN), a subcortical network, an uncertain network, the ventral attention network (VAN) and visual network. The FPN, CON and DMN are defined by both Fair and Cole, although Cole report more ROIs belonging to each of them. The function of these have been discussed above. The SMN is involved in multiple senses including pain, temperature, touch and limb position [9]. The SAN is involved in identifying what signals in the brain are most relevant for further processing and focus [91, 134]. The DAN and VAN are involved in attention, mediating voluntary focus on locations or features and detecting unexplained stimuli [152]. Auditory and visual networks are, as their names suggest, involved in listening

and vision. The subcortical network was a coherent functional network structure found within the subcortical areas of the brain, and the unknown network was a network their analysis found which could not be identified as any previous known functional network.

In contrast to Fair and Cole, the automated anatomical labelling (AAL) atlas [149] is a system of ROIs based solely on anatomy, not function, comprised of 116 regions. Figure 4.7 shows a mid sectional slice of the AAL atlas, where each colour represents a different anatomical region. We extracted time series for each region as the mean of all the voxel BOLD time series that fall within each region.

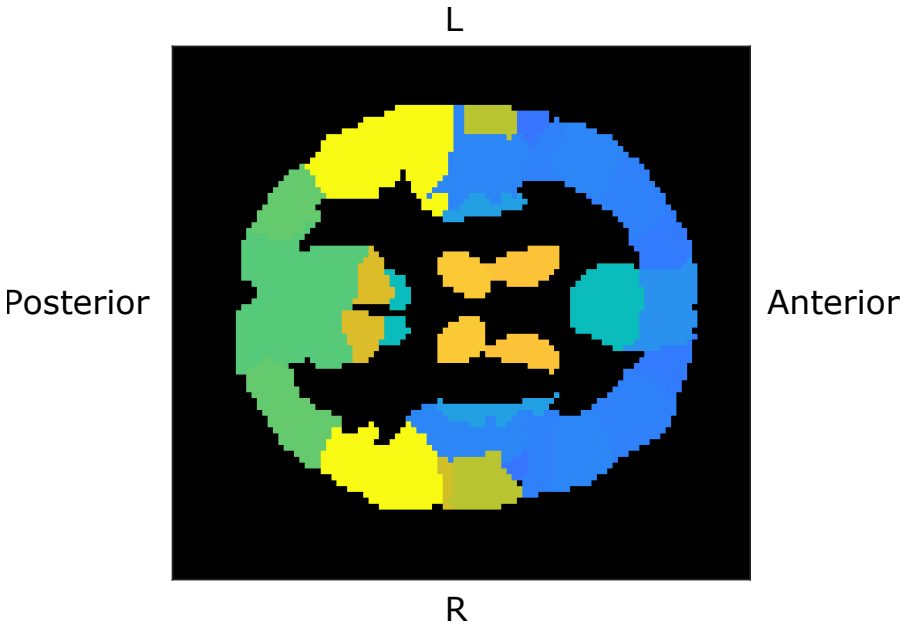


Figure 4.7: Axial slice of the AAL atlas produced from AAL mask in MATLAB; different colours represent different anatomical regions.

4.5 Network analysis methods

Functional connectivity (FC) is a statistical measure of how correlated the BOLD signal of two regions is, with a higher FC inferring greater communication. We measured FC using Pearson's correlation coefficient, pairwise for every ROI, resulting in a connectivity matrix for each participant. This matrix can be analysed as an adjacency matrix, describing a functional brain network. We kept the weightings of the network, rather than applying a threshold and transforming it into a binary network, because choosing a threshold has been shown to be ambiguous [33, 66]. One way to visualize how Pearson's correlation captures FC is to plot the BOLD values of one time series against another. Figure 4.8 shows four BOLD against BOLD plots along with their corresponding Pearson correlation coefficients, demonstrating different FC.

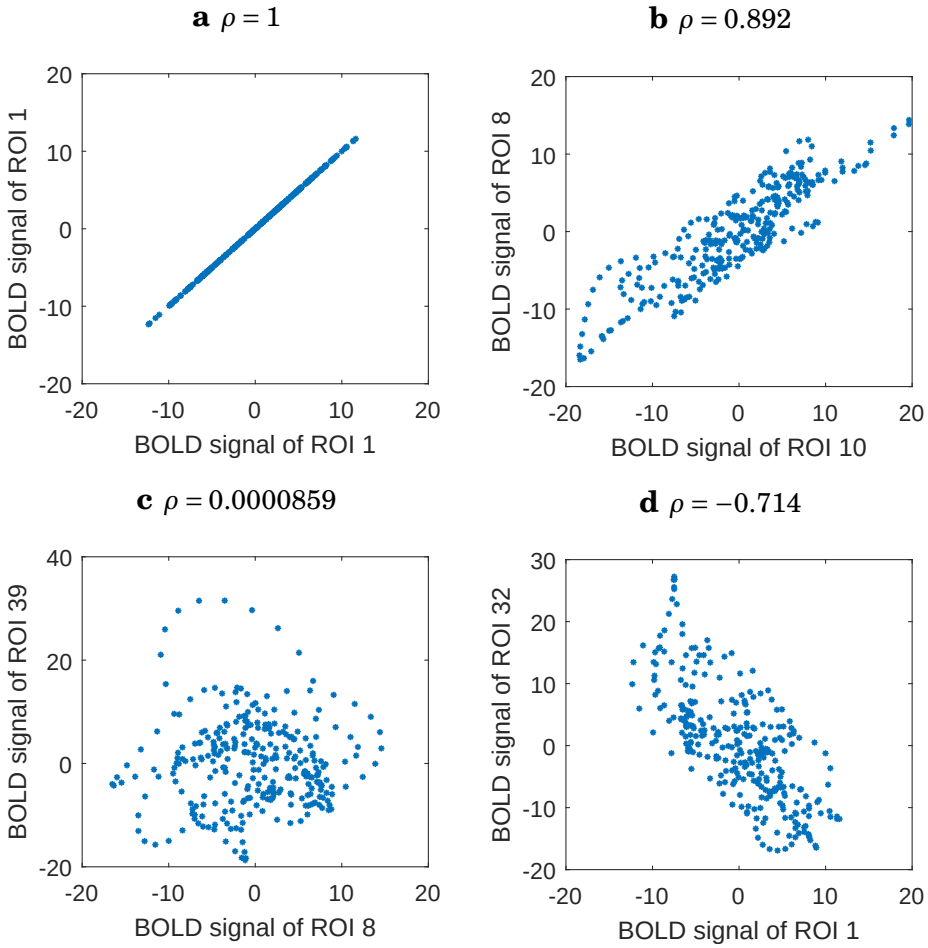


Figure 4.8: Visual representation of Pearson's correlation between a selection of different AAL ROIs. This is for the HC participant with a scanning time of 4.53 minutes, i.e. 300 volumes.

We applied the Fisher transform to the matrix to improve normality, defined as

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \operatorname{arctanh}(r), \tag{4.19}$$

where r is the input correlation coefficient and z is the transformed value. Correlation coefficients are bound between -1 and 1; if a number of sample correlations are near these bounds their distribution would be skewed. Applying the Fisher transformation removes these bounds and preserves variance across the correlation coefficient range; resulting in approximately more normal values.

Correlation coefficients can be both positive and negative. However, in fMRI there is little consensus as to the interpretation and how to handle negative FC weights [58]. This review

additionally searched resting-state studies and found a mixture of how they handled negative weights; with 57% reporting no or insufficient information on what they did, 21% discarded negative weights, 9% took the absolute value, 8% kept them and 5% did some other transformation. Due to this uncertainty we take the absolute value of the correlation as the FC measure, and argue that two strongly negatively correlated regions are still strongly functionally connected. Of note, for one part of our network analysis we do look at negative weights; see Modularity prior in Section 4.5.1.

4.5.1 Network measures

We looked at the following network properties to compare between the two populations. These network measures were part of the exploratory analysis; apart from basal ganglia connectivity, which was hypothesis driven.

Clustering coefficient

If two strongly connected neighbours of a node are themselves likely to be strongly connected to each other, the network is said to have a high clustering coefficient. This coefficient, C, is calculated as the average of all the local clustering coefficients, c_i, across all N nodes given by

$$C = \frac{1}{N}\sum_i c_i. \tag{4.20}$$

Defining the local clustering coefficient, c_i, is straightforward in binary networks, defined as the fraction of immediate neighbours also connected. Three extensions to weighted networks were explored by [97], with three differently defined local clustering coefficients. Onnela [109] define their local clustering coefficient as

$$c_{i,O} = \frac{1}{k_i(k_i-1)}\sum_{j,k}\left(\hat{w_{ij}}\hat{w_{jk}}\hat{w_{ik}}\right)^{\frac{1}{3}}. \tag{4.21}$$

The coefficient defined by Barrat [8] is

$$\begin{aligned} c_{i,B} &= \frac{1}{s_i(k_i-1)}\sum_{j,k}\frac{w_{ij}+w_{ik}}{2}a_{ij}a_{jk}a_{ik}, \\ s_i &= \sum_j w_{ij}, \end{aligned} \tag{4.22}$$

and the formulation by Zhang [167] is given by

$$c_{i,Z} = \frac{\sum_{j,k}\hat{w_{ij}}\hat{w_{jk}}\hat{w_{ik}}}{\left(\sum_k \hat{w_{ik}}\right)^2 - \sum_k\left(\hat{w_{ik}}^2\right)}, \tag{4.23}$$

where w_{ij} is the strength of the connection between nodes i and j, $\hat{w_{ij}} = w_{ij}/\max(w)$, k_i is the number of edges connected to node i, and $a_{ij} = 1$ if there is a connection between nodes i and j, otherwise it is 0. We only used Onnela's clustering coefficient because Barrat's one is used for networks which are not fully connected which is not the case with our network and Zhang's was developed in the context of gene expression networks, not functional brain networks.

Characteristic path length

The path length is defined as the ability for information to flow, directly or indirectly, between any two nodes; with shorter path lengths meaning flow is easier. The characteristic path length of a network is the average of all the path lengths in the network. The distance of a path between two regions should be short if the two nodes are more functionally connected. Since higher FC implies greater connectivity, distance was defined as the inverse of the connection strength, i.e.

$$d_{ij} = \frac{1}{w_{ij}}, \tag{4.24}$$

where w_{ij} is the FC between nodes i and j and d_{ij} is the new defined path length between them. Characteristic path length is the average of all the pair-wise distances, given by

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}. \tag{4.25}$$

Weighted small-world propensity

A network which displays small world properties simultaneously has large clustering coefficient and short characteristic path length [154]. However this formulation is for binary, unweighted networks. We used an extension to weighted networks [97], which calculates the weighted small-world propensity (w-SWP) by seeing how an input network deviates from two null models with equal weighting distribution and number of nodes. The null models are a lattice network, which have the highest clustering coefficients; and a random network, which have the shortest characteristic path lengths.

The fractional deviation of the clustering coefficient and characteristic path length from the null models are given respectively by

$$\Delta C = \frac{C_{\text{latt}} - C_{\text{obs}}}{C_{\text{latt}} - C_{\text{rand}}} \tag{4.26}$$

and

$$\Delta L = \frac{L_{\text{obs}} - L_{\text{rand}}}{L_{\text{latt}} - L_{\text{rand}}}, \tag{4.27}$$

where C_{latt} and L_{latt} are the clustering coefficient and average path length of the lattice network respectively, C_{rand} and L_{rand} are for the random network and C_{obs} and L_{obs} are for the input network of interest. The w-SWP is calculated from these as given by

$$\sigma = 1 - \sqrt{\frac{(\Delta C)^2 + (\Delta L)^2}{2}}. \tag{4.28}$$

Some networks can sometimes exceed lattice and random networks for the clustering coefficient and characteristic path length respectively. Therefore these fractional deviations (ΔC and ΔL) are capped between 0 and 1.

Modularity

The modularity quantifies the extent to which a network can be subdivided into densely connected regions called communities [47]. By construction, the FC within communities tends to be large and that between different communities small. We looked at two different modularity measures; one which finds the modularity of an optimized segregation of the whole network, and another which measures the modularity of the network already subdivided, based on prior labelling into distinct functional groups. These will be referred to as modularity optimized and modularity prior respectively. We calculated modularity optimized using code from the Brain Connectivity Toolbox [125], whose modularity optimized algorithm is based on prior work from [103, 104, 120].

Modularity prior was based on algorithms developed by [126], whose focus was on network measurements specifically for weighted functional brain networks. Modularity prior is given by

$$Q = \frac{1}{v} \sum_{ij} (w_{ij} - e_{ij}) \delta_{M_i M_j} \tag{4.29}$$

where $v = \sum_{ij} w_{ij}$, $e_{ij} = \frac{s_i s_j}{v}$, $s_i = \sum_j w_{ij}$ and $s_j = \sum_i w_{ij}$. For these calculations w_{ij} is the FC value between ROIs i and j and $\delta_{M_i M_j}$ is equal to 1 if ROIs i and j belong in the same module, otherwise it is 0.

An alteration to this modularity measure takes into account negative weights as well [126]. A high modularity for this measurements means a network displays high positive within module connectivity and high negative between module connectivity. Their reasoning is that positive FC values represent similar activation patterns, hence warranting placement in the same module

and negative FC values represent distinct, antiphase coupling, hence placement in different modules. The calculation for the modularity from positive weight contributions is given by

$$Q^+ = \frac{1}{v^+}\sum_{ij}(w_{ij}^+ - e_{ij}^+)\delta_{M_iM_j} \tag{4.30}$$

where $v^+ = \sum_{ij} w_{ij}^+$, $e_{ij}^+ = \frac{s_i^+ s_j^+}{v^+}$, $s_i^+ = \sum_j w_{ij}^+$ and $s_j^+ = \sum_i w_{ij}^+$. Here $w_{ij}^+ = |w_{ij}|$ if $w_{ij} > 0$ otherwise $w_{ij}^+ = 0$. Similarly, for the negative weights contributions the modularity is calculated as

$$Q^- = -\frac{1}{v^-}\sum_{ij}(w_{ij}^- - e_{ij}^-)\delta_{M_iM_j} \tag{4.31}$$

where $v^- = \sum_{ij} w_{ij}^-$, $e_{ij}^- = \frac{s_i^- s_j^-}{v^-}$, $s_i^- = \sum_j w_{ij}^-$ and $s_j^+ = \sum_i w_{ij}^+$. Here $w_{ij}^- = |w_{ij}|$ if $w_{ij} < 0$ otherwise $w_{ij}^- = 0$. The total modularity, taking into account positive and negative weights and based on prior module labelling, is calculated as

$$Q^* = Q^+ + \frac{v^-}{v^+ + v^-}Q^-. \tag{4.32}$$

Basal ganglia connectivity

The basal ganglia, as defined in the AAL, is comprised of six regions. The left and right Caudate, left and right Putamen, and left and right Pallidum. We define the basal inner connectivity as the average FC over the 15 region pairs within the basal ganglia. Basal outer connectivity was defined as the FC between a basal ganglia region and a region outside the basal ganglia, averaged over $6 \times 110 = 660$ such region pairs. We defined outer-outer connectivity as the average of the FC over the remaining 5,995 region pairs outside the basal ganglia.

4.5.2 Individual basal ganglia region analysis and single threshold permutation correction

We performed t-tests of the FC on each of the 15 region-region pairs within the basal ganglia separately, i.e. FC between left Putamen and right Putamen, FC between left Putamen and Left Pallidum etc.

With the individual basal ganglia connectivity analysis there is now a multiple comparisons problem, as it is comprised of 15 simultaneous statistical tests. Therefore, some form of multiple comparisons correction is needed. The multiple comparison correction we used is based on a permutation procedure [106], set out below. Rather than the p-value, the correction uses the t-statistic from each test, which is a standardized measure of the difference of means between

two populations. We performed this particular correction because it is non-parametric (so no parameters need fitting or altering) and it does not over-zealously correct every positive test (i.e. it is more 'forgiving').

The procedure is as follows:

1: From the MS labels and HC labels, randomly assign them to each participant, effectively shuffling who is labelled as MS or HC, whilst keeping the size of the two groups the same.

2: Perform basal ganglia analysis, measuring t-statistic for each region-region pair. A positive t-statistic means the FC between those two regions is greater in HC than MS. A negative t-statistic is the inverse (i.e. MS > HC).

3: Across all the region-region pair test results (total of 15) record the highest value (labelled $T_{\max}$) and the lowest value (labelled $T_{\min}$).

4: Re-shuffle the participant labels as in the first step, and repeat the analysis. This results in distributions of the maximum and minimum t-statistics across random shuffling of labels.

5: Corrected p-value of a region-region pair for HC>MS is the proportion of values in the permuted $T_{\max}$ distribution greater than the actual t-statistic for that region-region pair. For MS>HC it is the proportion of values in the permuted $T_{\min}$ distribution less than the actual t-statistic.

4.6 Network analysis results

In summary, the network measures we looked at were:

- Clustering Coefficient.
- Characteristic Path Length.
- Weighted small-world propensity (w-SWP).
- Modularity Prior - Absolute Weights (Fair and Cole only).
- Modularity Prior - Positive and negative Weights (Fair and Cole only).
- Modularity Optimized.
- Number of modules found from Modularity Optimize.
- Basal Inner. (AAL only).

- Basal Outer. (AAL only).
- Outer-outer. (AAL only).

We measured each network property for each participant and compared the properties between HC and MS using un-paired t-tests. This results in a p-value for each network property, representing similarity between the two groups. Table 4.2 shows the p-values for each network property, and for each relevant system of ROIs they fall within. Using a significance level of 0.05 we found only one significant result across all the systems of ROIs and network measures, that of basal inner mean connectivity with a p-value of 0.0177 (HC > MS).

	ROIs		
Network Measure	**Fair**	**Cole**	**AAL**
Clustering Coefficient	0.876	0.388	0.506
Characteristic Path Length	0.894	0.543	0.928
Weighted Small-world propensity	0.592	0.842	0.967
Modularity Prior (Absolute)	0.0964	0.934	-
Modularity Prior (Positive and Negative)	0.179	0.729	-
Modularity Optimise	0.914	0.329	0.493
Number of optimised Modules Found	0.566	0.324	0.476
Basal Inner Mean connectivity	-	-	0.0177
Basal Outer Mean connectivity	-	-	0.850
Outer-Outer Mean connectivity	-	-	0.961

Table 4.2: P-values from unpaired t-tests for different network measures, and different systems of ROIs, between MS and HC populations (0.01-0.08 Hz bandpass filter). Network measure not being applicable to the system of ROIs is indicated by a dash (-).

4.6.1 Individual basal ganglia region results

The average inner FC within the basal ganglia was found to be significantly different between HC and MS participants, prompting a more detailed investigation. Therefore we performed region by region FC analysis within the basal ganglia. Figure 4.9 shows a heatmap of p-values for each of the respective pairs, Figure 4.9a is all the p-values for which the FC of the region pairs was found greater in HC than MS, and vice versa for Figure 4.9b (i.e. MS greater than HC).

Only one region-regions pair's FC was found to be greater in MS than in HC, that between the left and the right Caudate. However this was not significant with a p-value close to one. For HC greater than MS four of the p-values were below 0.05: between left Putamen and right Putamen, left Putamen and right Pallidum, right Putamen and right Pallidum, and between left Pallidum and right Pallidum.

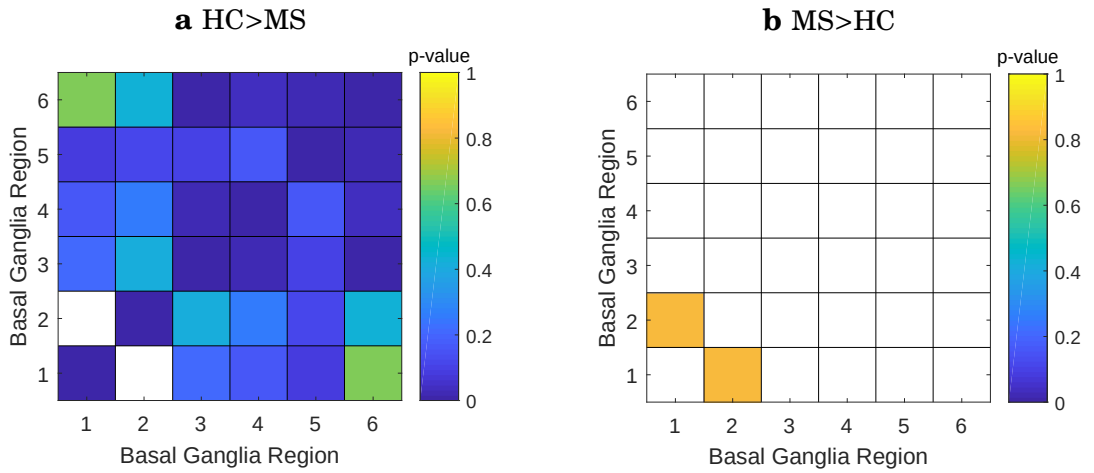


Figure 4.9: Individual Basal Ganglia regions FC t-test colourmap. KEY: 1. Left Caudate, 2. Right Caudate, 3. Left Putamen, 4. Right Putamen, 5. Left Pallidum, 6. Right Pallidum.

Significant results were found within the basal ganglia, but with 15 simultaneous statistical tests a multiple comparisons correction is needed. We performed the correction as set out in Section 4.5.2. Figure 4.10 shows results of the multiple corrections, displaying the distributions of the randomly shuffled $T_{\max}$ and $T_{\min}$ values in blue, along with the actual t-statistic values as red flags. Figure 4.10a shows the distribution of $T_{\max}$ values, significant region-region pairs would have their actual value (red flag in figure) greater than the distribution (i.e. to the right). Most of the actual values are firmly within the distribution, with the two furthest to the right having corrected p-values of 0.0639 and 0.0767. Figure 4.10b shows the distribution of $T_{\min}$ values, where significant region-region pairs would be to the left of the distribution. No value here is remotely significant, with the lowest corrected p-value being 0.941.

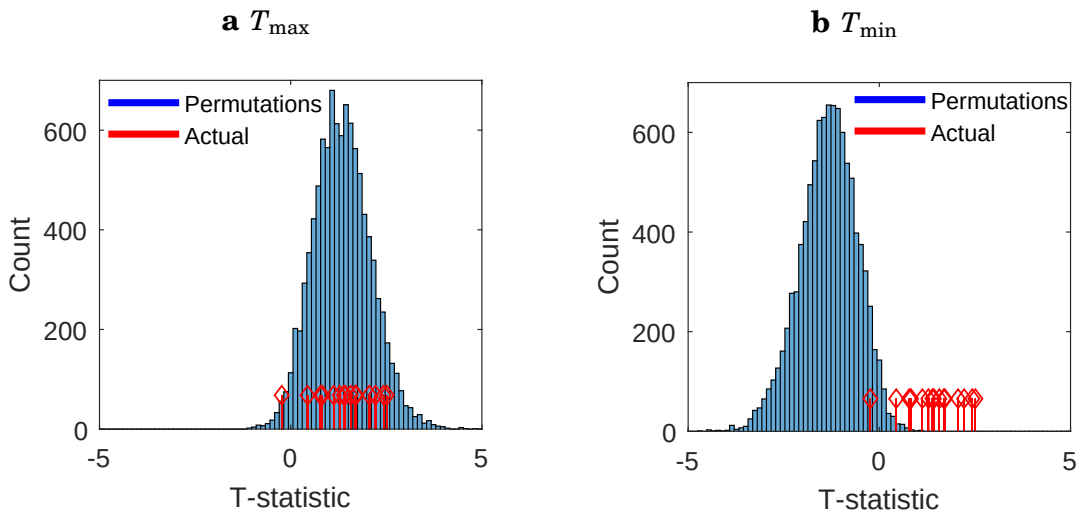


Figure 4.10: Distributions of random permutations of participant labelling used for multiple corrections. Distribution values are the maximum (**a**) or minimum (**b**) t-statistic measured across all 15 basal ganglia statistical comparisons for each permutation. Red flags plotted on top are the actual t-statistic values measured.

4.7 Correlations with clinical markers

The systematic review of Chapter 3 highlighted how a majority of the studies found performed correlations of a FC metric and some form of clinical marker, i.e. a test score (see Section 3.3.8). Our collaborators measured a number of test scores and clinical markers which we used to correlate with our network measures. A significant correlation between a network measure and one of the test scores or clinical markers could show evidence of compensatory neuroplasticity, even though none of the network measures themselves were significantly different between MS and HC. We performed this analysis using Pearson's correlation coefficient of each score against each network measure, for MS and HC separately. This results in a coefficient value and a p-value for each correlation.

The Clinical markers measured were: Age, CSF, D2 Error, D2 Speed, FSS, GM, Interference, Naming, T2, T2 RT, T3, T3 RT, Trail A, Trail B, WM and Words; a total of 16 markers.

CSF, GM and WM are cerebral spinal fluid, grey matter and white matter volume respectively. FSS is the fatigue severity scale, a questionnaire to evaluate the impact of fatigue. T2 and T3 are the tasks carried out in the scanner as part of the task based portion of the study; where T2 RT and T3 RT are the their respective reaction times. T2 was measuring intrinsic alertness and T3 was extrinsic. D2 Error and D2 Speed are part of a cognitive test called D2, measuring their error and speed performing this respectively. Naming, Interference and Words are part of the

Stroop test, used to assess cognitive interference, and can be used to measure other multiple cognitive functions [128]. Trail A and Trail B are hand written attention tests. These later tests (D2, Stroop, Trail A, Trial B) were carried out before and after the scan.

All this analysis was carried out in an exploratory manner; apart from correlations with the FSS, which is part of the hypothesis involving fatigue. Of note, other clinical measures displaying significant correlation with FC could still be interrupted back to fatigue, though it would be a weaker, indirect relationship.

For HC with Fair ROIs; out of 112 correlations (16 markers multiplied by 7 network measures), 5 had p-values less than 0.05, with the lowest value being 0.0147. With Cole ROIs 9 out of 112 (16 markers multiplied by 7 network measures) had a p-value less than 0.05, with the lowest being 0.00432, and for AAL ROIs 12 out of 128 correlations (16 markers multiplied by 8 network measures) passed the significance threshold, the lowest being 0.000117. For MS with Fair ROIs 10 out of 112 correlations passed significance, the lowest value being 0.0102; with Cole ROIs 6 out of 112 were lower than 0.05, the lowest being 0.0136. Finally, with AAL ROIs 7 out of 128 were significant, the lowest being 0.00897. CRICBristol recruited MS participants because of their fatigue. However none of the significant correlations were with FSS scores. We observed no clear tendency in these results, though some were lower than 0.05. However, none of these correlations are likely to pass multiple comparisons corrections, and those that did would require further interpretation to infer what physiological meaning they may have.

4.8 Preprocessing protocol alterations

The preprocessing protocol can affect group discrimination, as well as signal-noise separation and test-retest reliability [136]. This inspired us to look at how altering some of the preprocessing used might affect the significance of any differences found in network properties between the two populations. We looked at altering the nuisance regression signals and at the bandpass filter threshold values.

4.8.1 Nuisance regression alteration

Two of the studies found in the systematic review of Chapter 3 [60, 170] included the first time derivative of the nuisance signals in their regression. We did the same, with the results in Table 4.3. Comparing to the results with no first time derivative included, Table 4.2, no non-significant previous value has turned significant. However the single previous significant difference found, that of the basal inner, has had its p-value increase from 0.0177 to 0.0799, which is now considered non-significant.

	ROIs		
Network Measure	**Fair**	**Cole**	**AAL**
Clustering Coefficient	0.904	0.426	0.909
Characterstic Path Length	0.777	0.801	0.788
Weighted Small-world propensity	0.133	0.622	0.828
Modularity Prior (Absolute)	0.0901	0.974	-
Modularity Prior (Positive and Negative)	0.0996	0.505	-
Modularity Optimise	0.904	0.325	0.708
Number of optimised Modules Found	0.285	0.683	0.628
Basal Inner Mean connectivity	-	-	0.0799
Basal Outer Mean connectivity	-	-	0.769
Outer-Outer Mean connectivity	-	-	0.878

Table 4.3: P-values from unpaired t-tests for different network measures, and different systems of ROIs, between MS and HC populations (0.01-0.08 Hz bandpass filter). Including 1st time derivative of nuisance signals to regressors (compare to results in Table 4.2). Network measure not being applicable to the system of ROIs is indicated by a dash (-).

4.8.2 Bandpass parameter space

For all the previous analysis (results in Section 4.6) we used a fixed bandpass filter of 0.01-0.08 Hz. Previous studies have looked at how different frequencies contribute to FC [25, 160]. We therefore looked at how bandpass filter values affect the significance of differences between HC and MS over all the previous network metrics. We repeated the analysis using different values for the bandpass filter thresholds and recorded the p-value for each respective value. We used a range of no highpass to 0.05 Hz for the high pass cut off, and 0.06 to no low-pass for the low-pass cut off value. This resulted in a heat-map of p-values over the bandpass parameter space, representing significance of difference between HC and MS. The darker blue values represent there being a greater significance of difference between the populations, and lighter yellow values are for less significant differences. This shows significance, but not whether the measure is showing the value being greater in MS, or in HC. We plotted the p-values against their corresponding difference of the mean in measurement (mean HC minus mean MS) to check against a scenario in which a majority of the bandpass space is significant (i.e. mostly blue with majority of values less than 0.05), but a mix of significant values being MS greater than HC and vice versa.

Figure 4.11 shows the results of the bandpass space analysis for the clustering coefficient. Figure 4.11a is for Fair ROIs and shows that a majority of the p-values over the parameter space are large (mostly yellow). The red dots superimposed on this image are the different bandpass parameter values we found from the review in Chapter 3. Two of the values were outside the range we explored and are placed outside at approximately the location they correspond to. One of these (top right) was for a study designing a classifier and the other (bottom left) was focusing

on low frequency signals. These values are the same for all the bandpass space figures, and are only shown in this particular case for reference. Figure 4.11c is for Cole ROIs and Figure 4.11e for AAL ROIs; for both of these systems of ROIs there is a mix of significance across the bandpass space, though no areas are significantly dark. Figures 4.11b, 4.11d and 4.11f all show an even spread of p-values for MS greater than HC and vice versa for Fair, Cole and AAL respectively. This signifies no significant difference of this measure between HC and MS.

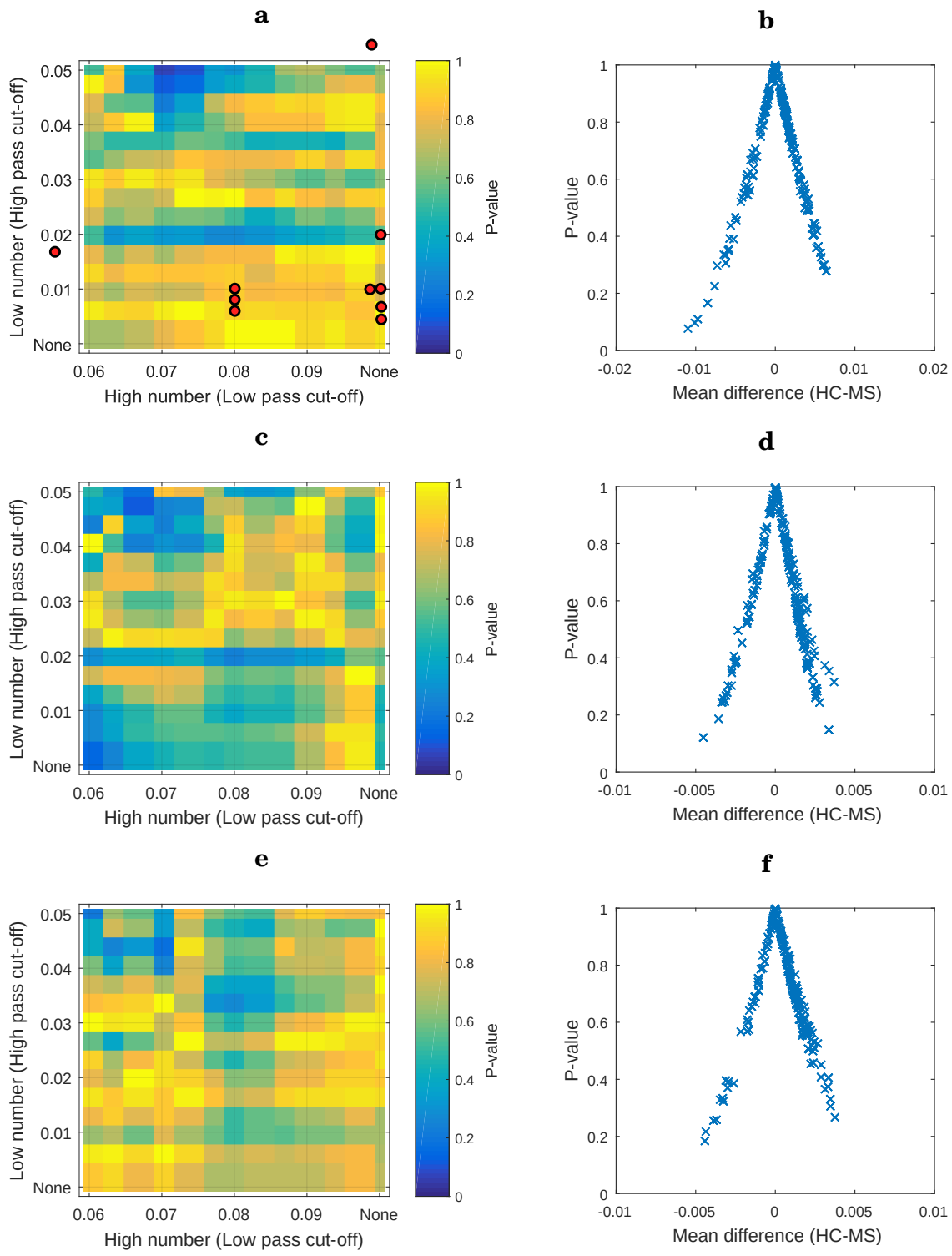


Figure 4.11: Bandpass parameter space results for the clustering coefficient. **a**,**c**,**e** Heatmap of p-values over a bandpass parameter set of values for Fair, Cole and AAL respectively. Red dots of **a** are parameter values found from the review in Chapter 3. **b**,**d**,**f** All points in heatmaps p-values plotted against the difference of mean that produced the respective p-value, for Fair, Cole and AAL respectively.

Figure 4.12 is the bandpass space analysis results for weighted small-world propensity. Figure 4.12a for Fair ROIs shows significance varies greatly. Figure 4.12b shows that a portion of the values are significant and that all significant bandpass values show w-SWP being greater in MS than in HC. When referring back to Figure 4.12a all the significant values (dark blues) are from the top right of the bandpass space, not a region commonly used (as found from the systematic review). Figure 4.12c is for Cole ROIs and shows the significance between HC and MS varies greatly over the bandpass parameter space, with some parts (slice across the middle) falling below 0.05. Figure 4.12d shows all the significant values show the MS having greater w-SWP than HC. However, large parts of the space are still non-significant reducing confidence in this being a significant result. For AAL ROIs, Figure 4.12e shows varying significance across the parameter space, with the greatest significance at the top left corner. Figure 4.12f shows all the highly significant values all show MS having greater w-SWP than HC. However it also shows that most of the values are still not significant.

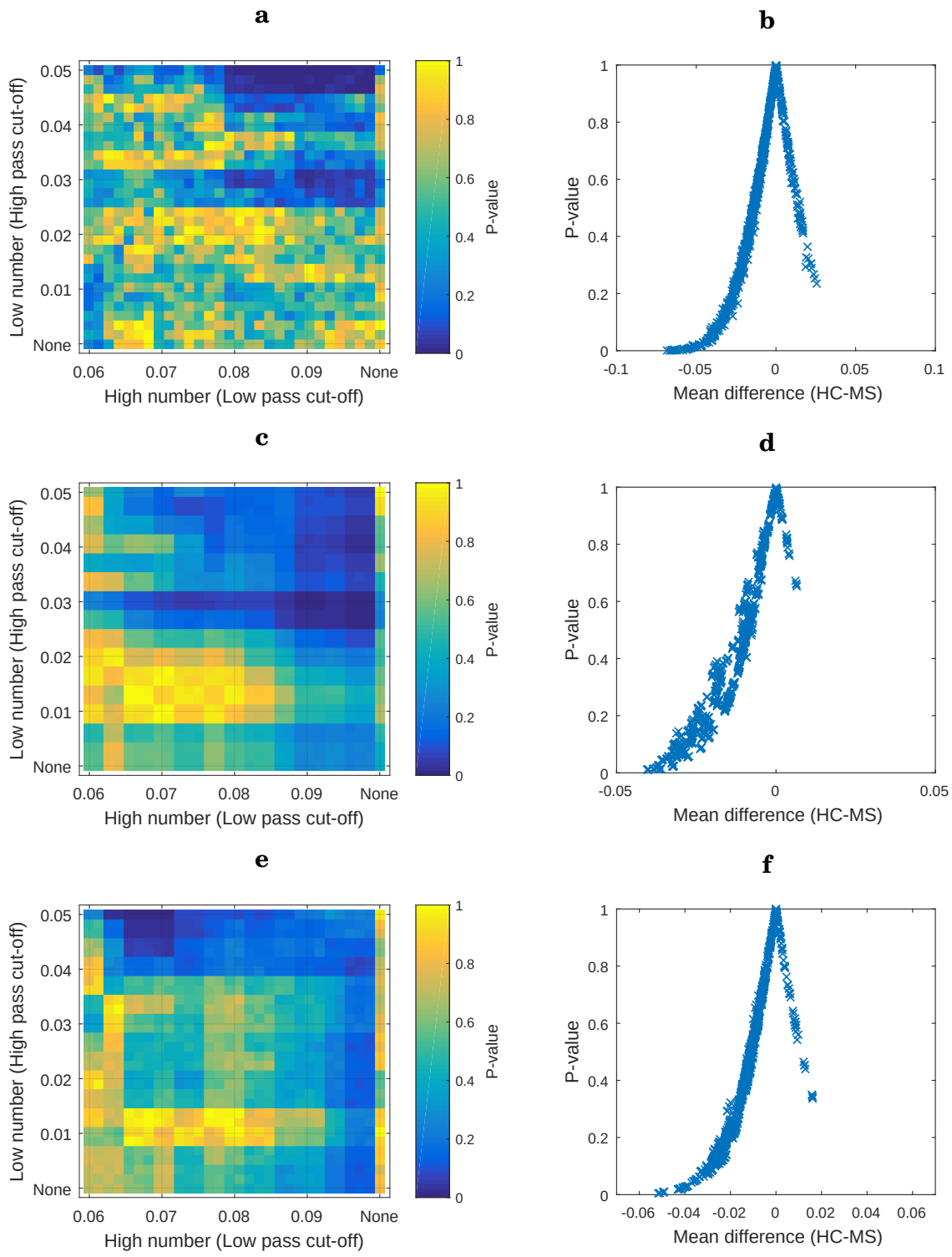


Figure 4.12: Bandpass parameter space results for the weighted small-world propensity. **a**,**c**,**e** Heatmap of p-values over a bandpass parameter set of values for Fair, Cole and AAL respectively. **b**,**d**,**f** All points in heatmaps p-values plotted against the difference of mean that produced the respective p-value, for Fair, Cole and AAL respectively.

Figure 4.13 is for basal inner, with Figure 4.13a showing the whole bandpass space to be dark blue, implying significance. Figure 4.13b shows that all the values are showing HC having greater basal inner connectivity than MS, with all the values clustered at the bottom right. This figure indicates that, although not all the values are below 0.05, a large proportion are.

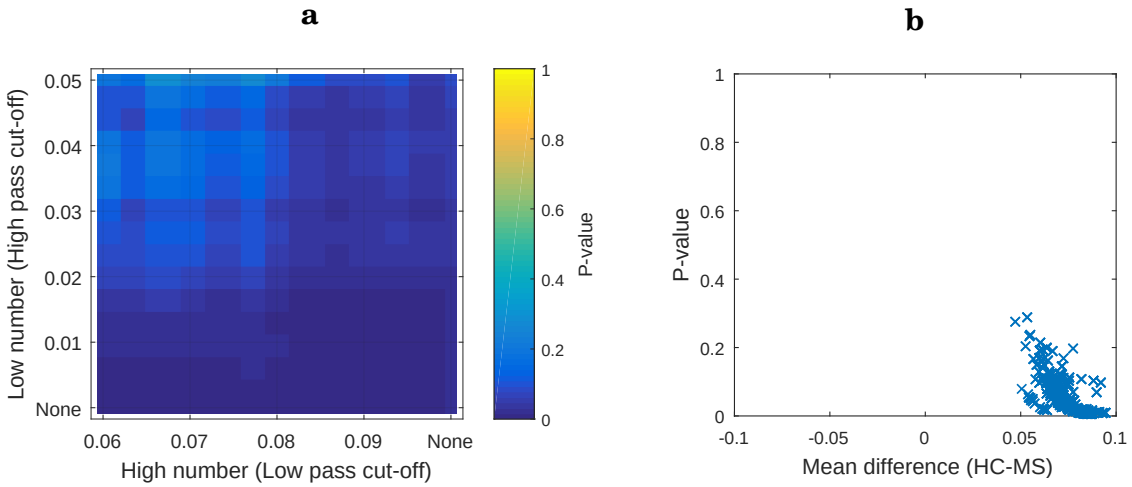


Figure 4.13: Bandpass parameter space of basal inner measurement (AAL). **a** Heatmap of p-values over a bandpass parameter set of values. **b** All points in heatmaps p-values plotted against the difference of mean that produced the respective p-value.

We explained the results of the bandpass space analysis for the clustering coefficient, weighted small-world propensity and basal inner only. All the other networks measures resulted in no significant results, and qualitatively looked the same as for the clustering coefficient, i.e. mostly insignificant p-values corresponding to a mix of MS greater than HC and vice versa.

4.9 Discussion

We found no significant difference between MS and HC with any of the network measures we looked at. This could be because the cohort of MS brains studied have already successfully undergone neuroplasticity to compensate for the damage caused.

Disruption to the basal ganglia has been widely reported to cause fatigue, which the MS participants recruited report as a common symptom of theirs. We found the average of all the basal ganglia's inner FC connections to be significantly different between HC and MS participants. This initially appears as good evidence as to a possible cause of the MS participants fatigue. However this significance was not preserved when analysing the individual connections and correcting for multiple comparisons. Additionally the mean inner FC of the basal ganglia was not significantly different when adding the first time derivative of the nuisance signals to regress,

reducing the confidence that there is a disruption to this part of the brain in the MS participants. A better ROI parcellation of the basal ganglia, representing a finer more course grain volume definition, may elucidate if there is any disruption. A further analysis we could perform to help validate these results is dynamic causal modelling (DCM) [49]. DCM is a method used to estimate directed adjacency matrices, i.e. whether a region directly influences another region. This method is suited for analysing a small number of regions, ideal for the six of the basal ganglia. A model is hypothesized describing how regions directly affect other regions, and DCM calculates how well the data fits the model. For the case here, based on our knowledge of what links we believe to be affected by MS, we could put forward two models of how the basal ganglia connects. One would be how the basal ganglia is thought to work in healthy populations, i.e. a "healthy model", and the other how we believe the connectivity of the basal ganglia may have been affected by MS, i.e. a "disrupted model". These models would both be run with HC and MS participants and if it is shown that the HC population fits the "healthy model" better than MS, and the MS population fits the "disrupted model" better than HC then this could be evidence of disruption to those particular links. This would require the models to be formulated before the DCM analysis could be run.

At the frequency band we chose, 0.01-0.08 Hz, Small-world propensity did not significantly differ between HC and MS. A study showed Small-world features in fMRI data to be most salient at lower frequencies (0.03-0.06 Hz) [5]. However when looking at the bandpass space results none of the system of ROIs showed significance of w-SWP at these frequencies, though they did at other frequencies; with Fair ROIs we showed significant differences at approximately 0.05-0.09 Hz, with Cole ROIs at 0.025-0.09 Hz and with AAL ROIs at around 0.05-0.065 Hz. The discrepancies in these values, and the fact that w-SWP for most of the bandpass space was still not significant reduces the confidence in there being any difference between the function of the brains of HC and MS in terms of small world saliency.

We chose to use Pearson's correlation coefficient to measure FC. This is an arbitrary choice and other measures do exist such as Spearman's correlation or mutual information, as found in the review of Chapter 3. However, the different measurements qualitatively measure the same thing: correlation between time series. Another possible way to measure FC would be by using partial correlation, measuring the correlation between two ROI time series whilst controlling for the correlation between all the rest of the ROIs.

The time series of each ROI are the average of the time series of a group of voxels, which are themselves the average measure of the BOLD signal of many brain cells. Each ROI time series could have a different variance depending on the voxels they were sampled from, especially for the AAL atlas whose different anatomical regions are different sizes and therefore comprised of a

different number of voxels. Analysing the variance of each time series could shape a different interpretation of any results that may be produced.

We found no evidence for what may be causing the reported fatigue of the MS participants, and with no network measure significantly differing between MS and HC, no mechanism can be put forward to its possible cause. Furthermore none of the measures significantly correlated with the fatigue severity scale.

CHAPTER 5

A Method of Characterizing Dynamic Functional Connectivity in fMRI

As discussed in Chapters 3 and 4, functional connectivity (FC) is a statistical measure of correlation between two regions of the brain, with higher FC inferring greater communication between the regions. When previously calculating the FC between two BOLD time series we used the Pearson's correlation coefficient over the entire time series. This is termed static FC and is an average of the FC between two regions over that time. The brain is dynamic and averaging over the whole time series could miss out important dynamics [21].

In this chapter we are interested in how the FC between two regions changes over time, rather than what its FC over that whole time period is. This measurement is called dynamic FC (rather than static FC) and there are different methods to calculate dynamic FC, with a recent review [163] outlining seven of them. Instead of outputting a single FC value, as for static FC, dynamic FC returns a time series of the changing FC over time. This presents a problem in terms of how to analyse and characterize the connectivity, which although lots of methods exist to analyse time series [79] these measures are not immediately applicable to dynamic FC. We wanted to investigate dynamic FC, using the Kalman filter as a potential tool for future research. We chose the Kalman filter because they are ideally suited to infer underlying trends or properties in noisy data, ideal for fMRI. Kalman filters are predict and update algorithms used for smoothing and estimation of parameters in many diverse fields [166].

We are aware of of one study which used a Kalman filter for magnetoencephalography (MEG) data [148]. However, to our knowledge no such analysis has been performed on dynamic functional connectivity with fMRI.

5.1 Methods

The Kalman filter works by combining the output of an underlying model at a certain state with external data measured at that state. This process is carried out sequentially with each subsequent state. Figure 5.1 demonstrates how the Kalman filter works, where a state (red circle) is updated via its underlying model (to a blue circle), and combined with data (fuzzy circle) at that state to output a better estimate for that state. This new state can then be updated and combined with further data, and the process repeats.

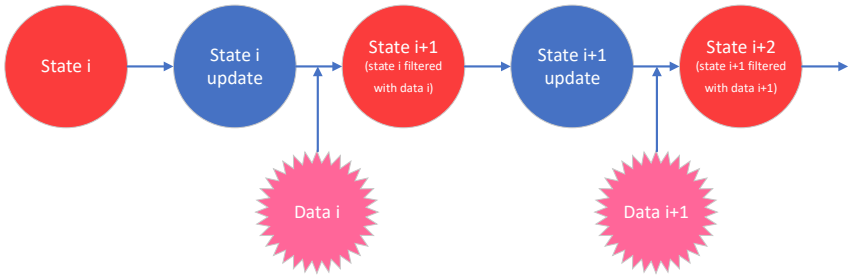


Figure 5.1: Representation overview of how a Kalman filter works. States (red circles) are updated via an underlying model (to a blue circle) and combined with data (fuzzy pink circles) for a better estimate of the state.

We apply the Kalman filter to fMRI via calculating dynamic FC, where the FC between two regions is the state. The system we are analysing here is 1-dimensional, that dimension being FC. We assume a simple underlying model where there is no change from one state to the next. This assumption is used because the data is from resting-state, i.e. no external influences, and this chapter is preliminary work. These methods could be extended to task based fMRI data, where the underlying model update would be correlated with the task.

We used a simple moving window method to calculate dynamic FC, as the Kalman Filter proposed here should be able to be applied to different methods of calculating dynamic FC.

Figure 5.2 shows how the moving window method works in calculating dynamic FC. Figure 5.2a shows two BOLD time series (blue) with a moving window (red). The two series are extracted from the Left Caudate and Right Caudate from a HC participant. The width of the window in this demonstration is 15 time to repetition (TR) units; TR is the sampling rate of fMRI, used as the time scale here. Pearson's correlation is performed between all the BOLD values within this window, and Fisher transformed. The window is moved along one time step and the calculation is

repeated. Figure 5.2b shows the resulting dynamic FC as the window is moved along, displaying how the connectivity between these two regions is dynamic over time.

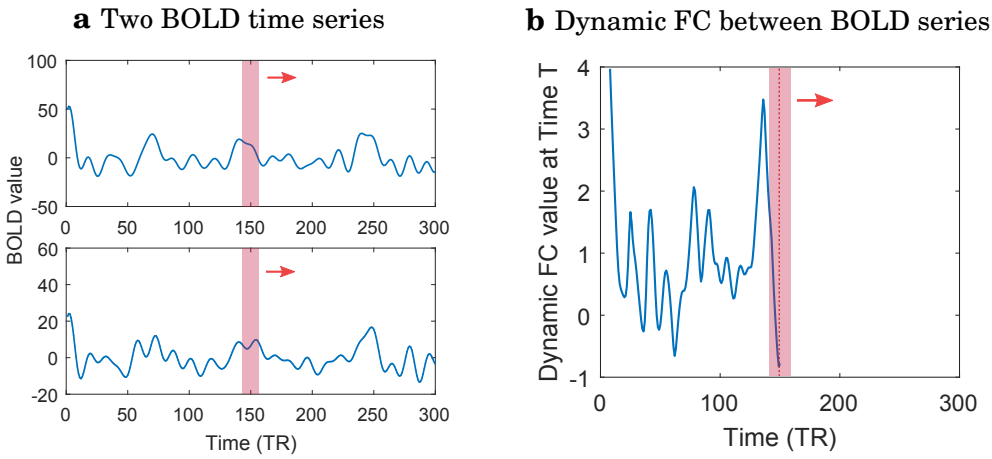


Figure 5.2: Moving window method for calculating dynamic FC. Moving window (red bar) moves left to right. **a** Two example time series to calculate dynamic FC of. Pearsons correlation is calculated within the moving window and Fisher z-transformed, with value recorded down as correlation at time halfway in the window. **b** Snapshot of Dynamic FC between example BOLD time series up to the moving window.

The Kalman filter is applied to the dynamic FC between two ROIs time series, for a single participant. Each moving window correlation value is the new data on which the filter updates. The filter starts with an initial guess of the FC, and with each iteration (as the time window scans across) updates this guess using each subsequent value from the dynamic correlation. The starting prior could be anything, and future work would need to be carried out to explore how this effects its performance. As we are neglecting the underlying model in this analysis, the Kalman update process demonstrated below is simpler than it would have been including an underlying model.

The Kalman filter updates using a Bayesian method, assuming all the data is normally distributed. Figure 5.3 shows a schematic to illustrate how the distributions of data and prior update to the posterior, displaying a distribution for an initial guess (prior in red), along with distributions for the data (blue) and the updated measurement for the value (posterior in yellow). After the measurement estimate has been updated with a step of the Kalman process, and since we have neglected the underlying model, the posterior distribution then becomes the prior of the next iteration.

The update of these distributions is done by looking at the means and variances of all the distributions involved. The mean and variance of the update measurement (posterior) is calculated using the Kalman gain:

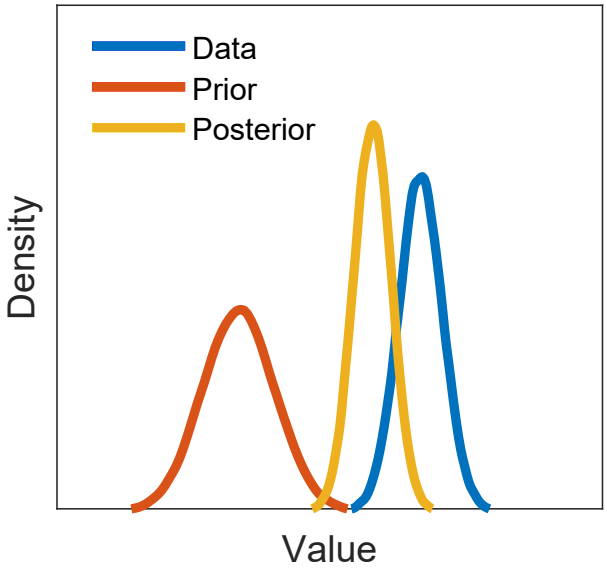


Figure 5.3: Graphical representation of Bayesian update of normal distributions.

$$\mathbf{K} = \frac{\sigma_P^2}{\sigma_P^2 + \sigma_D^2}, \tag{5.1}$$

which is used to define the posterior mean:

$$\mu_N = \mu_P + \mathbf{K}(\mu_D - \mu_P), \tag{5.2}$$

and the posterior variance:

$$\sigma_N^2 = (1 - \mathbf{K})^2 \, \sigma_P^2 + \mathbf{K}^2 \, \sigma_D^2, \tag{5.3}$$

where $\mathbf{K}$ is the Kalman gain; μ_N and σ_N are the updated mean (posterior mean) and updated standard deviation (posterior s.d), respectively; μ_P and σ_P are the starting mean (prior mean) and starting standard deviation (prior s.d), respectively; and μ_D and σ_D are the mean and standard deviation of the incoming information (data mean and data s.d), respectively.

The Kalman gain, $\mathbf{K}$, works such that if a new measurement has a very small variance there is greater certainty in its value, which draws the posterior mean μ_N closer to it. The inverse is true for a high variance, which has greater uncertainty and therefore the posterior mean μ_N will remain closer to the prior. As for the updated variance, each update qualitatively averages the variance of the prior and the data such that as long as both σ_D and σ_P are sufficiently small the

new variance σ_P will become smaller with each update.

Each subsequent measurement of the FC within each time window was measured using Pearson's correlation coefficient. This value is bound between -1 and 1 and is not normally distributed. This is why Fisher's z-transform was applied to each correlation measurement, which normalizes Pearson's correlation coefficient. The new Fisher z-transformed value becomes the mean of the data distribution. For its variance, Fisher's transform has the property that when applied to a bivariate normal distribution the standard error is simply a function of the number of data points used to calculate the correlation: $\frac{1}{\sqrt{N-3}}$, where N is the number of data points within each window. Figure 5.4 shows distributions of the BOLD values for two example ROIs, which appear normally distributed. This demonstrates that the correlation coefficient comes from a bivariate normal distribution, and that we are justified in assuming that the BOLD values are normally distributed. We therefore use Fisher's standard error as the value of σ_D in each update.

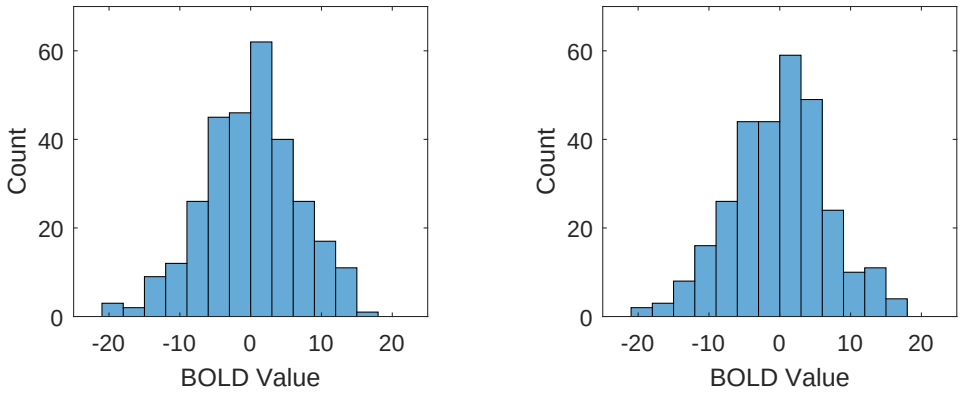


Figure 5.4: Distributions of the BOLD values for the two example ROI time series respectively.

5.2 Results

The Kalman filter applied to the FC between the same two example ROIs as before (i.e. left and right Caudate of a HC participant) is shown in Figure 5.5, displaying the dynamic correlation (blue) with the Kalman filtered correlation (red) along with the static correlation (dotted black line). This result was produced with a time window size of 10 TRs.

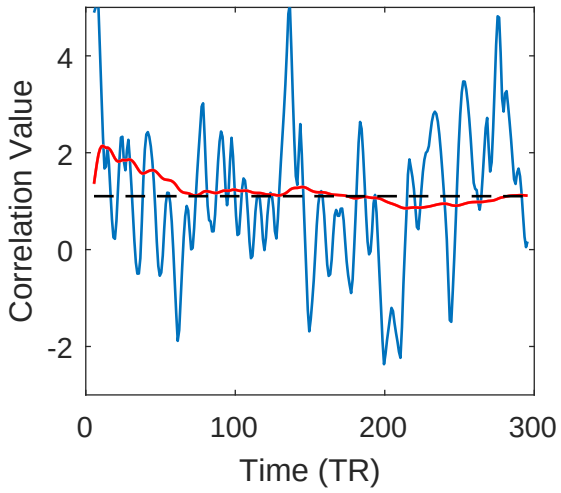


Figure 5.5: Dynamic correlation (blue) and Kalman output (red) with a window size of 10 TRs. Black dotted line is original FC calculation of entire time series at once.

The results for four different sizes of time window are shown in Figure 5.6. As for with a window size of 10 TR the filter works as a smoother, tracking the broad changes in FC over time.

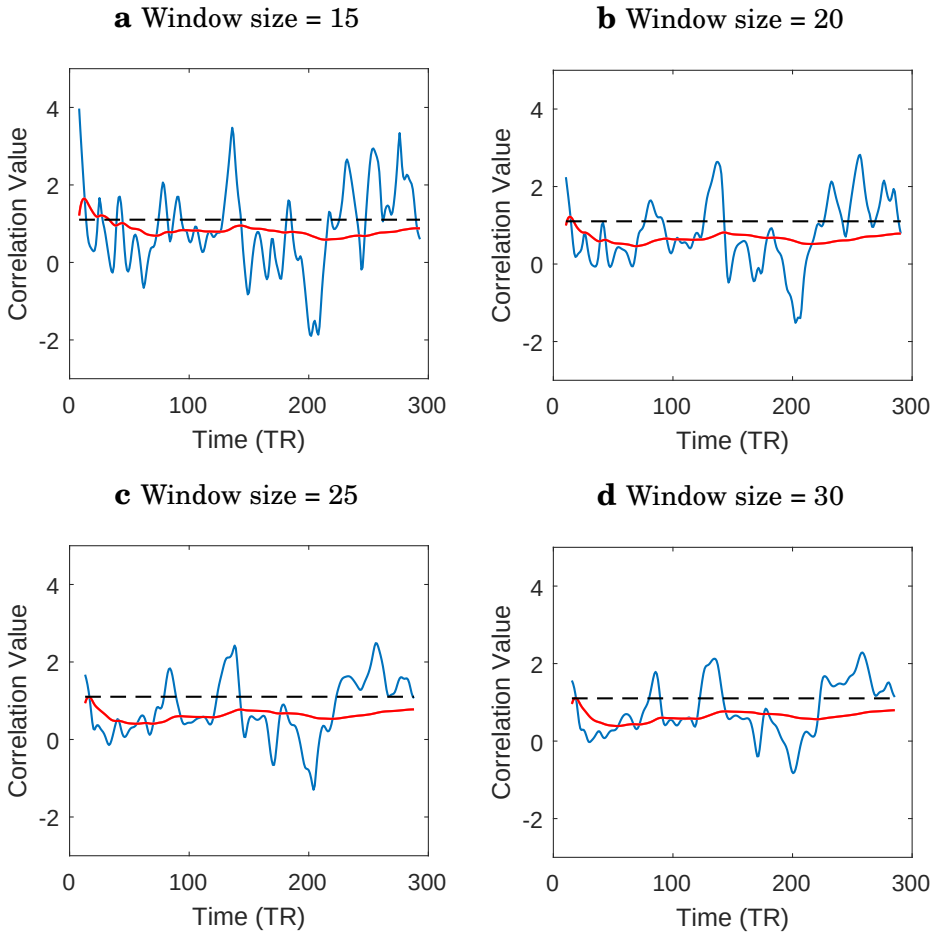


Figure 5.6: Dynamic correlation (blue) and Kalman output (red) plots with varying window size. Black dotted lines are original FC calculation of entire time series at once. Starting prior mean and variance (i.e. at Time 0) is the mean and variance of every point in the dynamic FC (blue line). Future work will be in exploring different starting values.

We tested the filter by measuring the value at the end of the Kalman process, i.e. the value produced by the Kalman filter at the end of the time series. We chose the end value as a statistic to explore because it has taken all the information over the course of the time series and may output possible insights into the filters performance. We explored how the window size affects the Kalman filter results, over a greater range of window sizes, by recording the mean and variance at the end of the Kalman process and plotting it against the window size which produced it. Figure 5.7a shows how the end mean is affected by varying window size, and Figure 5.7b shows how it affects the end variance. This shows how faster transient dynamics can affect the correlation measured, as seen with the decrease in the end FC with smaller time windows (valley around 0-50 TRs wide in Figure 5.7a). As for the variance, with greater window size there are less measurements, and so the Kalman filter has less information on which to converge.

However with smaller window sizes each measurement has more uncertainty even if there are more measurements. Figure 5.7b shows this interplay between having less measurements each with smaller variance and more measurements each with greater variance, reaching a minimum end variance with a window sizes of approximately half the full time series length. A desired behaviour may be to find an output value with the least uncertainty. Using the end Kalman value as an output would require using a window size half the entire time series length to minimize its variance. However this would not necessarily be the best choice as it could miss out on the faster transient dynamics.

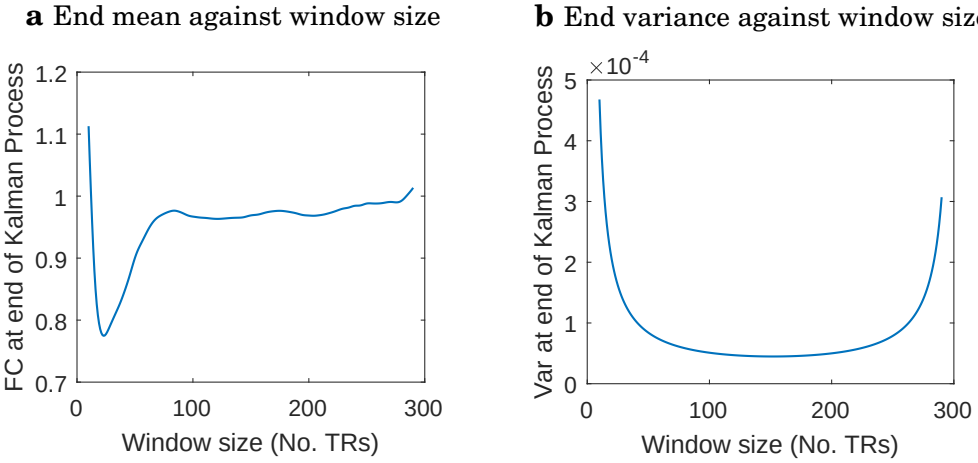


Figure 5.7: How the end of the Kalman process is affected by window size. Same single participant and ROIs as all other results in this chapter. **a** How mean of final Kalman estimate of the FC is affected by changing window size. **b** How the variance of the final Kalman estimate of the FC is affected by changing window size.

5.3 Discussion

This chapter has presented preliminary work on investigating conceptual tools to analyse dynamic FC via a Kalman Filter. It has not been used yet in analysing real world data and this chapter gives only an example between two time series. We show that the filter can track the broad changes in dynamic FC over time, and that considering faster dynamics with smaller window sizes can affect the results. Generalizing to other ROIs could throw up unpredictable behaviours that would require adjustments to the filter.

Future work on this would be to investigate how the starting mean and starting variance affect this Kalman process. Another area of research would be a more rigorous justification for choice of window size, as well as how a different window shape could affect the outcome. These questions have been researched in a recently published paper [95] applied to dynamic FC (though

not Kalman filters), and applying their findings to how the Kalman filter has been implemented here could be interesting.

Using a cumulative average of the dynamic FC would in this case have given very similar results. However the Kalman filter may out perform these in the case of when a signal is changing, as is the case in dynamic FC. Furthermore, for the example ROIs shown the Kalman filter acts qualitatively as a low-pass frequency filter. Moving averages and low-pass filters are an estimate of a measure part way to the current time, i.e. in the past. The Kalman filter would be an estimate at the current time, i.e. a live estimate, which may be a quality that means it could outperform a moving average. Future research could find the cases in which the Kalman filter proposed here can outperform simple cumulative or moving averages, and as the filter is a well established analysis technique plenty of tools exist that could be used to further analyse its results.

As Kalman Filters can be used for tracking of variables, one possible use of it in this case could be in real time estimation of FC, if MRI technology improved such that the time to capture BOLD images and process them was faster. For the same reason, the Kalman filter could be applied to fMRI measured when a participant is performing a task, rather than resting-state data. This is because a task happens in time, meaning the filter could better track the changing FC as a task progresses.

A key feature of the Kalman filter is the underlying model, which we have neglected. An underlying model with the Kalman filter could be used on task based studies, where the model follows the task. For resting-state data, the underlying model is not so clear; but one based on physiology could greatly improve its performance. Furthermore, with an underlying model, and more thorough development of the filter, new predictive methods could be established as this is a common use of the Kalman filter.

CHAPTER 6

FINAL CONCLUSIONS

The work in this thesis shows the breadth of areas applicable to network analysis. The two specific examples we explored were to a gene regulatory network (GRN) and to large scale functional brain networks. We showed that by abstracting complex biological systems into networks we can study simplified elements of them and produce useful insights into their workings.

Our aim for Chapter 2 was to modify a previously proposed GRN, describing interactions regulating pluripotency of mouse embryonic stem cells (mESCs), to include genes shown experimentally to be differentially expressed between two different culture media, Serum+LIF and 2i+LIF. Also, as all previous computational models of pluripotency GRNs only considered one of the inhibitors present in 2i+LIF (i.e. PD), we also aimed at including as a GRN input the other inhibitor Chiron. A new GRN was derived using published sequencing data, and a stochastic differential equation model derived to describe the dynamics of genes involved in response to the two drugs. Previous models were fitted to data generated with a tagged cell line recently shown to cause an incorrect measurement of the of the key pluripotency gene Nanog, due to high stability of the fluorescent tag. Instead, we fitted our model on data generated with a new and more reliable mESC line. We achieved our aims and demonstrated, combining numerical simulations of the model and bifurcation analysis, that both inhibitors can affect the GRN dynamics and, ultimately, mESC pluripotency. We also showed that, although not essential to describe Nanog dynamics, each of the newly added genes and interactions in the GRN can significantly alter the system dynamics. Limitations of our model are that some of the parameters could not be directly fitted to data, thus some of the chosen values could be considered arbitrary. The results still bring useful insights into the broad qualitative, culture media dependent behaviour of pluripotency

genes in mESCs.

Future work could include performing further network analysis to predict different culture media protocols to achieve pluripotency maintenance, for example by activating/repressing other genes in the GRN, and extending the network to include genes and chemicals important for cell differentiation.

Chapter 3 was a systematic review, carried out to scope the existing literature on the kind of analysis used to find evidence of neuroplasticity in the brains of participants with multiple sclerosis (MS). We found that pure resting-state functional magnetic resonance imaging (fMRI) studies in MS were low in number, though high in quality. We identified six different methods used to report evidence of neuroplasticity in MS from resting-state fMRI, each having different advantages and disadvantages. All the studies in the review of Chapter 3 reported some evidence of neuroplasticity. However, the results were inconsistent between studies, possibly due to the heterogeneity of MS.

In Chapter 4 we recruited MS and healthy control (HC) participants to have a resting-state fMRI scan, with the aim of finding evidence of neuroplasticity in MS. We used network analysis techniques, comparing network properties between MS and HC. Only the average connectivity within the basal ganglia was found to be significantly different between MS and HC, being greater in HC. However, after looking at individual pairwise connections within the basal ganglia and performing a multiple comparisons correction, this significance disappeared. We found no evidence of neuroplasticity in our cohort of MS, maybe signifying that the MS participants have already undergone successful neural reorganization. It may be the case that the compensatory changes in the MS brains are too subtle to be picked up with fMRI, and it may also be that as MS is a highly heterogeneous disease [44], differences across individuals average out when one looks at the larger cohort.

Limitations to the work of this chapter are the length of the resting-state scans measured. Scanning times of 9-13 minutes or longer have been shown to greatly improve the reliability of fMRI results [11], which is greater than our study which scanned for 4.53 minutes. Another limitation is in the fMRI technology itself. It can only indirectly measure neuronal activity via the flow of oxygenated blood, called the blood oxygen level dependent (BOLD) signal. When a region is more active and requires more oxygenated blood, the BOLD signal does not instantly react. This delay is the BOLD response, and has a width of approximately 3 seconds, although clever techniques can make inferences in the 100ms time scale [53]. However, this is still an order of magnitude greater than the underlying process of neuronal activity, which takes place on a time scale of approximately 10ms [84]. Furthermore, each voxel that fMRI measures the BOLD signal

of can comprise thousands of individual neurons. Therefore, although fMRI's spatial resolution is greater than other brain imaging methods, it is still not high enough to resolve individual neuronal activity.

Future work would be to apply the analysis pipeline developed here to other existing data which already shows evidence of neuroplasticity. This would verify the analysis pipeline or highlight possible issues with it. Using a different analysis technique, as highlighted in the systematic review of Chapter 3, could elucidate neuroplasticity in the MS brains. For example, using ICA or a voxel based analysis. However, more compelling differences would ideally show up over a variety of methods, i.e. invariant across different analysis techniques. Another analysis technique we could use is dynamic causal modelling (DCM) [49]. This method infers directed adjacency matrices, i.e. what region causally influences another region, and is suited to a small number of regions. This would be ideal for the basal ganglia, being comprised of only six regions, and could inform on any causal differences in the basal ganglia between the brains of MS and HC.

The work carried out in Chapter 5 was in investigating the application of a Kalman filter to dynamic functional connectivity (FC). We demonstrate that our filter can track the broad changes of FC over time, acting qualitatively as a smoother. The filter was only applied to one case example, generalizing to others could result in unpredictable behaviours. Some details of the filter still need exploring, such as how the starting parameters (mean and variance) effect it. Additionally, the Kalman filter demonstrated here effectively neglected any underlying model. With the addition of this the filter could be used to track and estimate live FC between regions of the brain. Further work is required on how the filter could be meaningfully implemented in studies.

BIBLIOGRAPHY

[1] *Covidence systematic review software, veritas health innovation, melbourne, australia. available at www.covidence.org*.

[2] *National institute for health and care excellence (2014) multiple sclerosis in adults: management clinical guideline (cg186). available at: https://www.nice.org.uk/guidance/cg186*.

[3] E. ABRANCHES, E. BEKMAN, AND D. HENRIQUE, *Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of nanog expression*, PLOS ONE, 8 (2013), p. e59928.

[4] E. ABRANCHES, A. M. V. GUEDES, M. MORAVEC, H. MAAMAR, P. SVOBODA, A. RAJ, AND D. HENRIQUE, *Stochastic nanog fluctuations allow mouse embryonic stem cells to explore pluripotency*, Development, 141 (2014), pp. 2770–2779.

[5] S. ACHARD, R. SALVADOR, B. WHITCHER, J. SUCKLING, AND E. BULLMORE, *A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs*, The Journal of Neuroscience, 26 (2006), pp. 63–72.

[6] U. A. ALON, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall/CRC, 2006.

[7] J. BAIN, L. PLATER, M. ELLIOTT, N. SHPIRO, C. J. HASTIE, H. MCLAUCHLAN, I. KLEVERNIC, J. S. ARTHUR, D. R. ALESSI, AND P. COHEN, *The selectivity of protein kinase inhibitors: a further update*, Biochemical Journal, 408 (2007), pp. 297–315.

[8] A. BARRAT, M. BARTHÉLEMY, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *The architecture of complex weighted networks*, Proceedings of the National Academy of Sciences of the United States of America, 101 (2004), pp. 3747–3752.

[9] S. BENSMAIA, *Somatosensory System: Overview*, Springer New York, New York, NY, 2013, pp. 1–2.

[10] J. L. BERNAT, *Chronic disorders of consciousness*, The Lancet, 367 (2006), pp. 1181–1192.

[11] E. BIRN, R.M. AMD MOLLOY, R. PATRIAT, P. T., T. MEIER, G. KIRK, V. NAIR, M. MEYERAND, AND V. PRABHAKARAN, *The effect of scan length on the reliability of resting-state fmri connectivity estimates*, NeuroImage, 83 (2013), pp. 550–558.

[12] S. BONAVITA, A. GALLO, R. SACCO, M. DELLA CORTE, A. BISECCO, R. DOCIMO, L. LAVORGNA, D. CORBO, A. DI COSTANZO, F. TORTORA, M. CIRILLO, F. ESPOSITO, AND G. TEDESCHI, *Distributed changes in default-mode resting-state connectivity in multiple sclerosis*, Multiple Sclerosis Journal, 17 (2011), pp. 411–422.

[13] S. BONAVITA, R. SACCO, M. DELLA CORTE, S. ESPOSITO, M. SPARACO, A. D'AMBROSIO, R. DOCIMO, A. BISECCO, L. LAVORGNA, D. CORBO, S. CIRILLO, A. GALLO, F. ESPOSITO, AND G. TEDESCHI, *Computer-aided cognitive rehabilitation improves cognitive performances and induces brain functional connectivity changes in relapsing remitting multiple sclerosis patients: An exploratory study*, Journal of Neurology, 262 (2015), pp. 91–100.

[14] M. H. BONNET AND D. L. ARAND, *Hyperarousal and insomnia: state of the science*, Sleep Medicine Reviews, 14 (2010), pp. 9–15.

[15] T. J. BRALEY AND R. D. CHERVIN, *Fatigue in multiple sclerosis: mechanisms, evaluation, and treatment*, Sleep, 33 (2010), pp. 1061–1067.

[16] P. BRANAS, R. JORDAN, A. FRY-SMITH, A. BURLS, AND C. HYDE, *Treatments for fatigue in multiple sclerosis: a rapid and systematic review*, Health Technology Assessment, 4 (2000), pp. 1–61.

[17] P. BROWNE, D. CHANDRARATNA, C. ANGOOD, H. TREMLETT, C. BAKER, B. V. TAYLOR, AND A. J. THOMPSON, *Atlas of multiple sclerosis 2013: A growing global problem with widespread inequity*, Neurology, 83 (2014), pp. 1022–1024.

[18] E. BULLMORE AND O. SPORNS, *Complex brain networks: graph theoretical analysis of structural and functional systems*, Nature Reviews Neuroscience, 10 (2009), pp. 186–198.

[19] M. G. CARTER, C. A. STAGG, G. FALCO, T. YOSHIKAWA, U. C. BASSEY, K. AIBA, L. V. SHAROVA, N. SHAIK, AND M. S. KO, *An in situ hybridization-based screen for heterogeneously expressed genes in mouse es cells*, Gene Expression Patterns, 8 (2008), pp. 181–198.

[20] I. CHAMBERS, D. COLBY, M. ROBERTSON, J. NICHOLS, S. LEE, S. TWEEDIE, AND A. SMITH, *Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells*, Cell, 113 (2003), pp. 643–655.

[21] C. CHANG AND G. H. GLOVER, *Time-frequency dynamics of resting-state brain connectivity measured with fmri*, NeuroImage, 50 (2010), pp. 81–98.

[22] A. CHAUDHURI AND P. O. BEHAN, *Fatigue and basal ganglia*, Journal of the Neurological Sciences, 179 (2000), pp. 34–42.

[23] V. CHICKARMANE, C. TROEIN, U. A. NUBER, H. M. SAURO, AND C. PETERSON, *Transcriptional dynamics of the embryonic stem cell switch*, PLOS Computational Biology, 2 (2006), p. e123.

[24] M. W. COLE, J. R. REYNOLDS, J. D. POWER, G. REPOVS, A. ANTICEVIC, AND T. S. BRAVER, *Multi-task connectivity reveals flexible hubs for adaptive task control*, Nature Neuroscience, 16 (2013), pp. 1348–1355.

[25] D. CORDES, V. M. HAUGHTON, K. ARFANAKIS, J. D. CAREW, P. A. TURSKI, C. H. MORITZ, M. A. QUIGLEY, AND M. E. MEYERAND, *Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data*, American Journal of Neuroradiology, 22 (2001), pp. 1326–1333.

[26] D. CORONADO, M. GODET, P. Y. BOURILLOT, Y. TAPPONNIER, A. BERNAT, M. PETIT, M. AFANASSIEFF, S. MARKOSSIAN, A. MALASHICHEVA, R. IACONE, K. ANASTASSIADIS, AND P. SAVATIER, *A short g1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency*, Stem Cell Research, 10 (2013), pp. 118–131.

[27] J. T. COULL, J. ME., E. TD., F. CD., AND M. M., *Attentional effects of noradrenaline vary with arousal level: selective activation of thalamic pulvinar in humans*, NeuroImage, 22 (2004), pp. 315–322.

[28] J. CRUZ-GÓMEZ Á, N. VENTURA-CAMPOS, A. BELENGUER, C. ÁVILA, AND C. FORN, *The link between resting-state functional connectivity and cognition in ms patients*, Multiple Sclerosis Journal, 20 (2014), pp. 338–348.

[29] A. J. CRUZ GÓMEZ, N. VENTURA CAMPOS, A. BELENGUER, C. ÁVILA, AND C. FORN, *Regional brain atrophy and functional connectivity changes related to fatigue in multiple sclerosis*, PLOS ONE, 8 (2013), p. e77914.

[30] C. V. DANG, *Myc, metabolism, cell growth, and tumorigenesis*, Cold Spring Harbor Perspectives in Medicine, 3 (2013), p. a014217.

[31] L. DE GIGLIO, F. TONA, F. DE LUCA, N. PETSAS, L. PROSPERINI, V. BIANCHI, C. POZZILLI, AND P. PANTANO, *Multiple sclerosis: Changes in thalamic resting-state functional connectivity induced by a home-based cognitive rehabilitation program*, Radiology, 280 (2016), pp. 202–211.

[32] M. DE GRAAF, K. JAGER, C. ZOCCALI, AND F. DEKKER, *Matching, an appealing method to avoid confounding?*, Nephron Clinical Practice, 118 (2011), pp. 315–318.

[33] F. DE VICO FALLANI, J. RICHIARDI, M. CHAVEZ, AND S. ACHARD, *Graph analysis of functional brain networks: practical issues in translational neuroscience*, Philosophical Transactions of the Royal Society B: Biological Sciences, 369 (2014), p. 20130521.

[34] E. A. DESOUSA, R. H. ALBERT, AND B. KALMAN, *Cognitive impairments in multiple sclerosis: a review*, American Journal of Alzheimer's Disease and Other Dementias, 17 (2002), pp. 23–29.

[35] A. M. DOGONOWSKI, H. R. SIEBNER, P. SOELBERG SØRENSEN, O. B. PAULSON, T. B. DYRBY, M. BLINKENBERG, AND K. H. MADSEN, *Resting-state connectivity of pre-motor cortex reflects disability in multiple sclerosis*, Acta Neurologica Scandinavica, 128 (2013), pp. 328–335.

[36] N. U. F. DOSENBACH, D. A. FAIR, A. L. COHEN, B. L. SCHLAGGAR, AND S. E. PETERSEN, *A dual-networks architecture of top-down control*, Trends in Cognitive Sciences, 12 (2008), pp. 99–105.

[37] N. U. F. DOSENBACH, D. A. FAIR, F. M. MIEZIN, A. L. COHEN, K. K. WENGER, R. A. T. DOSENBACH, M. D. FOX, A. Z. SNYDER, J. L. VINCENT, M. E. RAICHLE, B. L. SCHLAGGAR, AND S. E. PETERSEN, *Distinct brain networks for adaptive and stable task control in humans*, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007), pp. 11073–11078.

[38] N. U. F. DOSENBACH, K. M. VISSCHER, E. D. PALMER, F. M. MIEZIN, K. K. WENGER, H. C. KANG, E. D. BURGUND, A. L. GRIMES, B. L. SCHLAGGAR, AND S. E. PETERSEN, *A core system for the implementation of task sets*, Neuron, 50 (2006), pp. 799–812.

[39] D. A. FADDAH, H. WANG, A. W. CHENG, Y. KATZ, Y. BUGANIM, AND R. JAENISCH, *Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse escs*, Cell Stem Cell, 13 (2013), pp. 23–29.

[40] L. FAGNOCCHI, A. CHERUBINI, H. HATSUDA, A. FASCIANI, S. MAZZOLENI, V. POLI, V. BERNO, R. L. ROSSI, R. REINBOLD, M. ENDELE, T. SCHROEDER, M. ROCCHIGIANI, Z. SZKARLAT, S. OLIVIERO, S. DALTON, AND A. ZIPPO, *A myc-driven self-reinforcing regulatory network maintains mouse embryonic stem cell identity*, Nature Communications, 7 (2016), p. 11903.

[41] A. FAIVRE, A. RICO, W. ZAARAOUI, L. CRESPY, F. REUTER, D. WYBRECHT, E. SOULIER, I. MALIKOVA, S. CONFORT-GOUNY, P. J. COZZONE, J. PELLETIER, J. P. RANJEVA,

AND B. AUDOIN, *Assessing brain connectivity at rest is clinically relevant in early multiple sclerosis*, Multiple Sclerosis Journal, 18 (2012), pp. 1251–1258.

[42] A. FAIVRE, E. ROBINET, M. GUYE, C. ROUSSEAU, A. MAAROUF, A. LE TROTER, W. ZAARAOUI, A. RICO, L. CRESPY, E. SOULIER, S. CONFORT-GOUNY, J. PELLETIER, S. ACHARD, J. P. RANJEVA, AND B. AUDOIN, *Depletion of brain functional connectivity enhancement leads to disability progression in multiple sclerosis: A longitudinal resting-state fmri study*, Multiple Sclerosis, 22 (2016), pp. 1695–1708.

[43] A. FAIVRE, A. AMD RICO, W. ZAARAOUI, F. REUTER, S. CONFORT-GOUNY, M. GUYE, J. PELLETIER, J.-P. RANJEVA, AND B. AUDOIN, *Brain functional plasticity at rest and during action in multiple sclerosis patients*, Journal of Clinical Neuroscience, 22 (2015), pp. 1438–1443.

[44] A. FEINSTEIN, P. O'CONNOR, T. GRAY, AND K. FEINSTEIN, *The effects of anxiety on psychiatric morbidity in patients with multiple sclerosis*, Multiple Sclerosis Journal, 5 (1999), pp. 323–326.

[45] M. FIDALGO, F. FAIOLA, C.-F. PEREIRA, J. DING, A. SAUNDERS, J. GINGOLD, C. SCHANIEL, I. R. LEMISCHKA, J. C. R. SILVA, AND J. WANG, *Zfp281 mediates nanog autorepression through recruitment of the nurd complex and inhibits somatic cell reprogramming*, Proceedings of the National Academy of Sciences of the United States of America, 109 (2012), pp. 16202–16207.

[46] A. FORNITO, A. ZALESKY, AND E. T. BULLMORE, *Fundamentals of brain network analysis*, Elsevier/Academic Press, 2016.

[47] S. FORTUNATO, *Community detection in graphs*, Physics Reports, 486 (2010), pp. 75–174.

[48] J. E. FREAL, G. H. KRAFT, AND J. K. CORYELL, *Symptomatic fatigue in multiple sclerosis*, Archives of Physical Medicine and Rehabilitation, 65 (1984), pp. 135–138.

[49] K. J. FRISTON, L. HARRISON, AND W. PENNY, *Dynamic causal modelling*, NeuroImage, 19 (2003), pp. 1273–1302.

[50] A. GALLO, F. ESPOSITO, R. SACCO, R. DOCIMO, A. BISECCO, M. D. CORTE, A. D'AMBROSIO, D. CORBO, N. ROSA, M. LANZA, S. CIRILLO, S. BONAVITA, AND G. TEDESCHI, *Visual resting-state network in relapsingremitting ms with and without previous optic neuritis*, Neurology, 79 (2012), pp. 1458–1465.

[51] O. L. GAMBOA, E. TAGLIAZUCCHI, F. VON WEGNER, A. JURCOANE, M. WAHL, H. LAUFS, AND U. ZIEMANN, *Working memory performance of early ms patients correlates inversely with modularity increases in resting state functional connectivity networks*, NeuroImage, 94 (2014), pp. 385–395.

[52] I. GLAUCHE, M. HERBERG, AND I. ROEDER, *Nanog variability and pluripotency regulation of embryonic stem cells - insights from a mathematical model analysis*, PLOS ONE, 5 (2010), p. e11238.

[53] G. H. GLOVER, *Overview of functional magnetic resonance imaging*, Neurosurgery Clinics of North America, 22 (2011), pp. 133–139.

[54] S. GODWIN, L. MARUCCI, N. MASUDA, J. THAI, AND C. VAN ANTWERPEN, *Functional connevtivity in multiple sclerosis (ms): a systematic review of resting state functional mri. prospero 2017 crd42017059861. available from: http://www.crd.york.ac.uk/prospero/display_record.php?id=crd42017059861*.

[55] S. GODWIN, D. WARD, E. PEDONE, M. HOMER, A. G. FLETCHER, AND L. MARUCCI, *An extended model for culture-dependent heterogenous gene expression and proliferation dynamics in mouse embryonic stem cells*, Nature Partner Journals Systems Biology and Applications, 3 (2017), p. 19.

[56] M. D. GREICIUS, B. KRASNOW, A. L. REISS, AND V. MENON, *Functional connectivity in the resting brain: a network analysis of the default mode hypothesis*, Proceedings of the National Academy of Sciences of the United States of America, 100 (2003), pp. 253–258.

[57] J. A. HACKETT AND M. A. SURANI, *Regulatory principles of pluripotency: from the ground state up*, Cell Stem Cell, 15 (2014), pp. 416–430.

[58] M. N. HALLQUIST AND F. G. HILLARY, *Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world*, Network Neuroscience, 3 (2019), pp. 1–26.

[59] S. HAQ, A. MICHAEL, M. ANDREUCCI, K. BHATTACHARYA, P. DOTTO, B. WALTERS, J. WOODGETT, H. KILTER, AND T. FORCE, *Stabilization of beta-catenin by a wnt-independent mechanism regulates cardiomyocyte growth*, Proceedings of the National Academy of Sciences of the United States of America, 100 (2003), pp. 4610–4615.

[60] D. J. HAWELLEK, J. F. HIPP, C. M. LEWIS, M. CORBETTA, AND A. K. ENGEL, *Increased functional connectivity indicates the severity of cognitive impairment in multiple sclerosis*, Proceedings of the National Academy of Sciences of the United States of America, 108 (2011), pp. 19066–19071.

[61] M. HERBERG, I. GLAUCHE, T. ZERJATKE, M. WINZI, F. BUCHHOLZ, AND I. ROEDER, *Dissecting mechanisms of mouse embryonic stem cells heterogeneity through a model-based analysis of transcription factor dynamics*, Journal of the Royal Society Interface, 13 (2016), p. 20160167.

[62] M. HERBERG, T. KALKAN, I. GLAUCHE, A. SMITH, AND I. ROEDER, *A model-based analysis of culture-dependent phenotypes of mescs*, PLOS ONE, 9 (2014), p. e92496.

[63] S. HERCULANO-HOUZEL, *The human brain in numbers: a linearly scaled-up primate brain*, Frontiers in Human Neuroscience, 3 (2009), p. 31.

[64] D. W. HUANG, B. T. SHERMAN, Q. TAN, J. KIR, D. LIU, D. BRYANT, Y. GUO, R. STEPHENS, M. W. BASELER, H. C. LANE, AND R. A. LEMPICKI, *David bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists*, Nucleic Acids Research, 35 (2007), pp. W169–W175.

[65] H. E. HULST, M. M. SCHOONHEIM, Q. VAN GEEST, B. M. UITDEHAAG, F. BARKHOF, AND J. J. GEURTS, *Memory impairment in multiple sclerosis: Relevance of hippocampal activation and hippocampal connectivity*, Multiple Sclerosis Journal, 21 (2015), pp. 1705–1712.

[66] M. JALILI, *Functional brain networks: Does the choice of dependency estimator and binarization method matter?*, Scientific Reports, 6 (2016), p. 29780.

[67] V. JANARDHAN AND R. BAKSHI, *Quality of life in patients with multiple sclerosis: the impact of fatigue and depression*, Journal of the Neurological Sciences, 205 (2002), pp. 51–58.

[68] A. L. JANSSEN, A. BOSTER, B. A. PATTERSON, A. ABDULJALIL, AND R. S. PRAKASH, *Resting-state functional connectivity in multiple sclerosis: an examination of group differences and individual differences*, Neuropsychologia, 51 (2013), pp. 2918–2929.

[69] L. J. JULIAN, L. VELLA, T. VOLLMER, O. HADJIMICHAEL, AND D. C. MOHR, *Employment in multiple sclerosis. exiting and re-entering the work force*, Journal of Neurology, 255 (2008), pp. 1354–1360.

[70] G. KOBELT, J. BERG, P. LINDGREN, S. FREDRIKSON, AND B. JÖNSSON, *Costs and quality of life of patients with multiple sclerosis in europe*, Journal of Neurology, Neurosurgery and Psychiatry, 77 (2006), pp. 918–926.

[71] A. A. KOLODZIEJCZYK, J. K. KIM, J. C. TSANG, T. ILICIC, J. HENRIKSSON, K. N. NATARAJAN, A. C. TUCK, X. GAO, M. BUHLER, P. LIU, J. C. MARIONI, AND S. A. TEICHMANN, *Single cell rna-sequencing of pluripotent states unlocks modular transcriptional variation*, Cell Stem Cell, 17 (2015), pp. 471–485.

[72] S. J. KUHL AND M. KUHL, *On the role of wnt/beta-catenin signaling in stem cells*, Biochimica et Biophysica Acta (BBA) - General Subjects, 1830 (2013), pp. 2297–2306.

[73] J. LANSLEY, D. MATAIX-COLS, M. GRAU, J. RADUA, AND J. SASTRE-GARRIGA, *Localized grey matter atrophy in multiple sclerosis: a meta-analysis of voxel-based morphometry studies and associations with functional disability*, Neuroscience and Biobehavioral Reviews, 37 (2013), pp. 819–830.

[74] H. LASSMANN, W. BRUCK, AND C. F. LUCCHINETTI, *The immunopathology of multiple sclerosis: an overview*, Brain Pathology, 17 (2007), pp. 210–218.

[75] Y. LIU, P. LIANG, Y. DUAN, J. HUANG, Z. REN, X. JIA, H. DONG, J. YE, F.-D. SHI, H. BUTZKUEVEN, AND K. LI, *Altered thalamic functional connectivity in multiple sclerosis*, European Journal of Radiology, 84 (2015), pp. 703–708.

[76] Y. LIU, P. LIANG, Y. DUAN, X. JIA, C. YU, M. ZHANG, F. WANG, M. ZHANG, H. DONG, J. YE, H. BUTZKUEVEN, AND K. LI, *Brain plasticity in relapsing-remitting multiple sclerosis: evidence from resting-state fmri*, Journal of the Neurological Sciences, 304 (2011), pp. 127–131.

[77] M. LOITFELDER, *Abnormalities of resting state functional connectivity are related to sustained attention deficits in ms*, PLOS ONE, 7 (2012), p. e42862.

[78] Y. LUO, C. L. LIM, J. NICHOLS, A. MARTINEZ-ARIAS, AND L. WERNISCH, *Cell signalling regulates dynamics of nanog distribution in embryonic stem cell populations*, Journal of the Royal Society Interface, 10 (2013), p. 20120525.

[79] D. MACHIWAL AND M. K. JHA, *Methods for Time Series Analysis. In: Hydrologic Time Series Analysis: Theory and Practice*, Springer, Dordrecht, 2012.

[80] I. S. MACKENZIE, S. V. MORANT, G. A. BLOOMFIELD, T. M. MACDONALD, AND J. O'RIORDAN, *Incidence and prevalence of multiple sclerosis in the uk 1990-2010: a descriptive study in the general practice research database*, Journal of Neurology, Neurosurgery, and Psychiatry, 85 (2014), pp. 76–84.

[81] C. MAINERO, P. PANTANO, F. CARAMIA, AND C. POZZILLI, *Brain reorganization during attention and memory tasks in multiple sclerosis: insights from functional mri studies*, Journal of the Neurological Sciences, 245 (2006), pp. 93–98.

[82] D. MALLEY, *Deluca, j. (ed.). (2005). fatigue as a window to the brain. london: The mit press*, Neuropsychological Rehabilitation, 16 (2006), pp. 597–599.

[83] H. MARKS, T. KALKAN, R. MENAFRA, S. DENISSOV, K. JONES, H. HOFEMEISTER, J. NICHOLS, A. KRANZ, A. F. STEWART, A. SMITH, AND H. G. STUNNENBERG, *The transcriptional and epigenomic foundations of ground state pluripotency*, Cell, 149 (2012), pp. 590–604.

[84] S. MAROM, *Neural timescales or lack thereof*, Progress in Neurobiology, 90 (2010), pp. 16–28.

[85] G. MARTELLO, T. SUGIMOTO, E. DIAMANTI, A. JOSHI, R. HANNAH, S. OHTSUKA, B. GOTTGENS, H. NIWA, AND A. SMITH, *Esrrb is a pivotal target of the gsk3/tcf3 axis regulating embryonic stem cell self-renewal*, Cell Stem Cell, 11 (2012), pp. 491–504.

[86] L. MARUCCI, *Nanog dynamics in mouse embryonic stem cells: Results from systems biology approaches*, Stem Cells International, 2017 (2017), p. 7160419.

[87] L. MARUCCI, E. PEDONE, U. DI VICINO, B. SANUY-ESCRIBANO, M. ISALAN, AND M. P. COSMA, *beta-catenin fluctuates in mouse escs and is essential for nanog-mediated reprogramming of somatic cells to pluripotency*, Cell Reports, 8 (2014), pp. 1686–1696.

[88] V. S. MATTAY, F. FERA, A. TESSITORE, A. R. HARIRI, S. DAS, J. H. CALLICOTT, AND D. R. WEINBERGER, *Neurophysiological correlates of age-related changes in human motor function*, Neurology, 58 (2002), pp. 630–635.

[89] W. I. MCDONALD, A. COMPSTON, G. EDAN, D. GOODKIN, H. P. HARTUNG, F. D. LUBLIN, H. F. MCFARLAND, D. W. PATY, C. H. POLMAN, S. C. REINGOLD, M. SANDBERG-WOLLHEIM, W. SIBLEY, A. THOMPSON, S. VAN DEN NOORT, B. Y. WEINSHENKER, AND J. S. WOLINSKY, *Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis*, Ann Neurol, 50 (2001), pp. 121–127.

[90] V. MENON, *Large-Scale Functional Brain Organization*, Academic Press, Waltham, 2015, pp. 449–459.

[91] V. MENON, *Salience Network*, Academic Press, Waltham, 2015, pp. 597–611.

[92] A. MINAGAR, M. H. BARNETT, R. H. B. BENEDICT, D. PELLETIER, I. PIRKO, M. A. SAHRAIAN, E. FROHMAN, AND R. ZIVADINOV, *The thalamus and multiple sclerosis: Modern views on pathologic, imaging, and clinical aspects*, Neurology, 80 (2013), pp. 210–219.

[93] G. R. MIRAMS, C. J. ARTHURS, M. O. BERNABEU, R. BORDAS, J. COOPER, A. CORRIAS, Y. DAVIT, S. J. DUNN, A. G. FLETCHER, D. G. HARVEY, M. E. MARSH, J. M. OSBORNE, P. PATHMANATHAN, J. PITT-FRANCIS, J. SOUTHERN, N. ZEMZEMI, AND D. J. GAVAGHAN, *Chaste: an open source c++ library for computational physiology and biology*, PLOS Computational Biology, 9 (2013), p. e1002970.

[94] Y. MIYANARI AND M. E. TORRES-PADILLA, *Control of ground-state pluripotency by allelic regulation of nanog*, Nature, 483 (2012), pp. 470–473.

[95] F. MOKHTARI, M. I. AKHLAGHI, S. L. SIMPSON, G. WU, AND P. J. LAURIENTI, *Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state*, NeuroImage, 189 (2019), pp. 655–666.

[96] F. M. MOTTAGHY, *Interfering with working memory in humans*, Neuroscience, 139 (2006), pp. 85–90.

[97] S. F. MULDOON, E. W. BRIDGEFORD, AND D. S. BASSETT, *Small-world propensity and weighted brain networks*, Nature Scientific Reports, 6 (2016), p. 22057.

[98] N. P. MULLIN, A. YATES, A. J. ROWE, B. NIJMEIJER, D. COLBY, P. N. BARLOW, M. D. WALKINSHAW, AND I. CHAMBERS, *The pluripotency rheostat nanog functions as a dimer*, Biochemical Journal, 411 (2008), pp. 227–231.

[99] S. MUNOZ DESCALZO, P. RUE, F. FAUNES, P. HAYWARD, L. M. JAKT, T. BALAYO, J. GARCIA-OJALVO, AND A. MARTINEZ ARIAS, *A competitive protein interaction network buffers oct4-mediated differentiation to promote pluripotency in embryonic stem cells*, Molecular Systems Biology, 9 (2013), p. 694.

[100] K. MURPHY, R. M. BIRN, AND P. A. BANDETTINI, *Resting-state fmri confounds and cleanup*, NeuroImage, 80 (2013), pp. 349–359.

[101] F. NAKAKI AND M. SAITOU, *Prdm14: a unique regulator for pluripotency and epigenetic reprogramming*, Trends in Biochemical Sciences, 39 (2014), pp. 289–298.

[102] P. NAVARRO, N. FESTUCCIA, D. COLBY, A. GAGLIARDI, N. P. MULLIN, W. ZHANG, V. KARWACKI-NEISIUS, R. OSORNO, D. KELLY, M. ROBERTSON, AND I. CHAMBERS, *Oct4/sox2-independent nanog autorepression modulates heterogeneous nanog gene expression in mouse es cells*, The Embo Journal, 31 (2012), pp. 4547–4562.

[103] M. E. NEWMAN, *Finding community structure in networks using the eigenvectors of matrices*, Physical Review E Covering Statistical, Nonlinear, Biological, and Soft Matter Physics, 74 (2006), p. 036104.

[104] M. E. NEWMAN, *Modularity and community structure in networks*, Proceedings of the National Academy of Sciences of the United States of America, 103 (2006), pp. 8577–8582.

[105] T. E. NICHOLS, S. DAS, S. B. EICKHOFF, A. C. EVANS, T. GLATARD, M. HANKE, N. KRIEGESKORTE, M. P. MILHAM, R. A. POLDRACK, J. B. POLINE, E. PROAL, B. THIRION, D. C. VAN ESSEN, T. WHITE, AND B. T. YEO, *Best practices in data analysis and sharing in neuroimaging using mri*, Nature Neuroscience, 20 (2017), pp. 299–303.

[106] T. E. NICHOLS AND A. P. HOLMES, *Nonparametric permutation tests for functional neuroimaging: a primer with examples*, Human Brain Mapping, 15 (2002), pp. 1–25.

[107] A. NISHIYAMA, L. XIN, A. A. SHAROV, M. THOMAS, G. MOWRER, E. MEYERS, Y. PIAO, S. MEHTA, S. YEE, Y. NAKATAKE, C. STAGG, L. SHAROVA, L. S. CORREA-CERRO, U. BASSEY, H. HOANG, E. KIM, R. TAPNIO, Y. QIAN, D. DUDEKULA, M. ZALZMAN, M. LI, G. FALCO, H. T. YANG, S. L. LEE, M. MONTI, I. STANGHELLINI, M. N. ISLAM, R. NAGARAJA, I. GOLDBERG, W. WANG, D. L. LONGO, D. SCHLESSINGER, AND M. S. KO, *Uncovering early response of gene regulatory networks in escs by systematic induction of transcription factors*, Cell Stem Cell, 5 (2009), pp. 420–433.

[108] C. W. NOONAN, D. WILLIAMSON, J. HENRY, R. INDIAN, S. LYNCH, J. NEUBERGER, R. SCHIFFER, J. TROTTIER, L. WAGNER, AND R. MARRIE, *The prevalence of multiple sclerosis in 3 us communities*, Preventing Chronic Disease, 7 (2010), p. A12.

[109] J. P. ONNELA, J. SARAMAKI, J. KERTESZ, AND K. KASKI, *Intensity and coherence of motifs in weighted complex networks*, Physical Review E Covering Statistical, Nonlinear, Biological, and Soft Matter Physics, 71 (2005), p. 065103.

[110] J. M. OSBORNE, A. G. FLETCHER, J. M. PITT-FRANCIS, P. K. MAINI, AND D. J. GAVAGHAN, *Comparing individual-based approaches to modelling the self-organization of multicellular tissues*, PLOS Computational Biology, 13 (2017), p. e1005387.

[111] L. PARISI, M. ROCCA, P. VALSASINA, L. PANICARI, F. MATTIOLI, AND M. FILIPPI, *Cognitive rehabilitation correlates with the functional connectivity of the anterior cingulate cortex in patients with multiple sclerosis*, Brain Imaging and Behavior, 8 (2014), pp. 387–393.

[112] W. PENNY, K. FRISTOL, J. ASHBURNER, S. KIEBEL, AND T. NICHOLS, eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press.

[113] B. PERIN, O. GODEFROY, S. FALL, AND G. DE MARCO, *Alertness in young healthy subjects: an fmri study of brain region interactivity enhanced by a warning signal*, Brain and Cognition, 72 (2010), pp. 271–281.

[114] N. PETSAS, V. TOMASSINI, N. FILIPPINI, E. SBARDELLA, F. TONA, M. C. PIATTELLA, C. POZZILLI, R. WISE, AND P. PANTANO, *Impaired functional connectivity unmasked by simple repetitive motor task in early relapsing-remitting multiple sclerosis*, Neurorehabilitation and Neural Repair, 29 (2015), pp. 557–565.

[115] C. H. POLMAN, S. C. REINGOLD, B. BANWELL, M. CLANET, J. A. COHEN, M. FILIPPI, K. FUJIHARA, E. HAVRDOVA, M. HUTCHINSON, L. KAPPOS, F. D. LUBLIN, X. MONTALBAN, P. O'CONNOR, M. SANDBERG-WOLLHEIM, A. J. THOMPSON, E. WAUBANT,

B. WEINSHENKER, AND J. S. WOLINSKY, *Diagnostic criteria for multiple sclerosis: 2010 revisions to the mcdonald criteria*, Annals of Neurology, 69 (2011), pp. 292–302.

[116] C. H. POLMAN, S. C. REINGOLD, G. EDAN, M. FILIPPI, H. P. HARTUNG, L. KAPPOS, F. D. LUBLIN, L. M. METZ, H. F. MCFARLAND, P. W. O'CONNOR, M. SANDBERG-WOLLHEIM, A. J. THOMPSON, B. G. WEINSHENKER, AND J. S. WOLINSKY, *Diagnostic criteria for multiple sclerosis: 2005 revisions to the "mcdonald criteria"*, Annals of Neurology, 58 (2005), pp. 840–846.

[117] C. M. POSER, D. W. PATY, L. SCHEINBERG, W. I. MCDONALD, F. A. DAVIS, G. C. EBERS, K. P. JOHNSON, W. A. SIBLEY, D. H. SILBERBERG, AND W. W. TOURTELLOTTE, *New diagnostic criteria for multiple sclerosis: guidelines for research protocols*, Annals of Neurology, 13 (1983), pp. 227–231.

[118] J. D. POWER, A. L. COHEN, S. M. NELSON, G. S. WIG, K. A. BARNES, J. A. CHURCH, A. C. VOGEL, T. O. LAUMANN, F. M. MIEZIN, B. L. SCHLAGGAR, AND S. E. PETERSEN, *Functional network organization of the human brain*, Neuron, 72 (2011), pp. 665–678.

[119] M. E. RAICHLE, A. M. MACLEOD, A. Z. SNYDER, W. J. POWERS, D. A. GUSNARD, AND G. L. SHULMAN, *A default mode of brain function*, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001), pp. 676–682.

[120] J. REICHARDT AND S. BORNHOLDT, *Statistical mechanics of community detection*, Physical Review E Covering Statistical, Nonlinear, Biological, and Soft Matter Physics, 74 (2006), p. 016110.

[121] G. RICCITELLI, M. A. ROCCA, E. PAGANI, V. MARTINELLI, M. RADAELLI, A. FALINI, G. COMI, AND M. FILIPPI, *Mapping regional grey and white matter atrophy in relapsing-remitting multiple sclerosis*, Multiple Sclerosis Journal, 18 (2012), pp. 1027–1037.

[122] J. RICHIARDI, M. GSCHWIND, S. SIMIONI, J. M. ANNONI, B. GRECO, P. HAGMANN, M. SCHLUEP, P. VUILLEUMIER, AND D. VAN DE VILLE, *Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity*, NeuroImage, 62 (2012), pp. 2021–2033.

[123] M. A. ROCCA, A. FALINI, B. COLOMBO, G. SCOTTI, G. COMI, AND M. FILIPPI, *Adaptive functional changes in the cerebral cortex of patients with nondisabling multiple sclerosis correlate with the extent of brain structural damage*, Annals of Neurology, 51 (2002), pp. 330–339.

[124] S. D. ROOSENDAAL, M. M. SCHOONHEIM, H. E. HULST, E. J. SANZ-ARIGITA, S. M. SMITH, J. J. G. GEURTS, AND F. BARKHOF, *Resting state networks change in clinically isolated syndrome*, Brain: a Journal of Neurology, 133 (2010), pp. 1612–1621.

[125] M. RUBINOV AND O. SPORNS, *Complex network measures of brain connectivity: uses and interpretations*, NeuroImage, 52 (2010), pp. 1059–1069.

[126] M. RUBINOV AND O. SPORNS, *Weight-conserving characterization of complex functional brain networks*, NeuroImage, 56 (2011), pp. 2068–2079.

[127] M. SANCHEZ-CASTILLO, D. RUAU, A. C. WILKINSON, F. S. NG, R. HANNAH, E. DIAMANTI, P. LOMBARD, N. K. WILSON, AND B. GOTTGENS, *Codex: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities*, Nucleic Acids Research, 43 (2015), pp. D1117–D1123.

[128] F. SCARPINA AND S. TAGINI, *The stroop color and word test*, Frontiers in psychology, 8 (2019), p. 557.

[129] N. D. SCHIFF, *Central thalamic contributions to arousal regulation and neurological disorders of consciousness*, Annals of the New York Academy of Sciences, 1129 (2008), pp. 105–118.

[130] M. P. SCHNETZ, L. HANDOKO, B. AKHTAR-ZAIDI, C. F. BARTELS, C. F. PEREIRA, A. G. FISHER, D. J. ADAMS, P. FLICEK, G. E. CRAWFORD, T. LAFRAMBOISE, P. TESAR, C. L. WEI, AND P. C. SCACHERI, *Chd7 targets active gene enhancer elements to modulate es cell-specific gene expression*, PLOS Genetics, 6 (2010), p. e1001023.

[131] M. M. SCHOONHEIM, J. J. G. GEURTS, O. T. WIEBENGA, J. C. DE MUNCK, C. H. POLMAN, C. J. STAM, F. BARKHOF, AND A. M. WINK, *Changes in functional network centrality underlie cognitive dysfunction and physical disability in multiple sclerosis*, Multiple Sclerosis Journal, 20 (2014), pp. 1058–1065.

[132] M. M. SCHOONHEIM, H. E. HULST, R. B. BRANDT, M. STRIK, A. M. WINK, B. M. UITDEHAAG, F. BARKHOF, AND J. J. GEURTS, *Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis*, Neurology, 84 (2015), pp. 776–783.

[133] M. M. SCHOONHEIM, H. E. HULST, D. LANDI, O. CICCARELLI, S. D. ROOSENDAAL, E. J. SANZ-ARIGITA, H. VRENKEN, C. H. POLMAN, C. J. STAM, F. BARKHOF, AND J. J. G. GEURTS, *Gender-related differences in functional connectivity in multiple sclerosis*, Multiple Sclerosis Journal, 18 (2012), pp. 164–173.

[134] W. W. SEELEY, V. MENON, A. F. SCHATZBERG, J. KELLER, G. H. GLOVER, H. KENNA, A. L. REISS, AND M. D. GREICIUS, *Dissociable intrinsic connectivity networks for salience processing and executive control*, The Journal of Neuroscience, 27 (2007), pp. 2349–2356.

[135] D. SEIXAS, J. PALACE, AND I. TRACEY, *Chronic pain disrupts the reward circuitry in multiple sclerosis*, European Journal of Neuroscience, 44 (2016), pp. 1928–1934.

[136] W. SHIRER, H. JIANG, C. PRICE, B. NG, AND M. GREICIUS, *Optimization of rs-fmri pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination*, NeuroImage, 117 (2015), pp. 67–79.

[137] A. M. SILLITO, J. CUDEIRO, AND H. E. JONES, *Always returning: feedback and sensory processing in visual cortex and thalamus*, Trends in Neurosciences, 29 (2006), pp. 307–316.

[138] J. SILVA AND A. SMITH, *Capturing pluripotency*, Cell, 132 (2008), pp. 532–536.

[139] Z. S. SINGER, J. YONG, J. TISCHLER, J. A. HACKETT, A. ALTINOK, M. A. SURANI, L. CAI, AND M. B. ELOWITZ, *Dynamic heterogeneity and dna methylation in embryonic stem cells*, Molecular Cell, 55 (2014), pp. 319–331.

[140] S. K. SINGH, A. MARISETTY, P. SATHYAN, M. KAGALWALA, Z. ZHAO, AND S. MAJUMDER, *Rest-mir-21-sox2 axis maintains pluripotency in e14tg2a.4 embryonic stem cells*, Stem Cell Research, 15 (2015), pp. 305–311.

[141] S. H. A. STROGATZ, *Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering*, Second edition. Boulder, CO : Westview Press, a member of the Perseus Books Group, 2015.

[142] W. STURM, A. DE SIMONE, B. J. KRAUSE, K. SPECHT, V. HESSELMANN, I. RADERMACHER, H. HERZOG, L. TELLMANN, H. W. MULLER-GARTNER, AND K. WILLMES, *Functional anatomy of intrinsic alertness: evidence for a fronto-parietal-thalamic-brainstem network in the right hemisphere*, Neuropsychologia, 37 (1999), pp. 797–805.

[143] W. STURM, F. LONGONI, B. FIMM, T. DIETRICH, S. WEIS, S. KEMNA, H. HERZOG, AND K. WILLMES, *Network for auditory intrinsic alertness: a pet study*, Neuropsychologia, 42 (2004), pp. 563–568.

[144] T. SWIRSKY-SACCHETTI, D. R. MITCHELL, J. SEWARD, C. GONZALES, F. LUBLIN, R. KNOBLER, AND H. L. FIELD, *Neuropsychological and structural brain lesions in multiple sclerosis: a regional analysis*, Neurology, 42 (1992), pp. 1291–1295.

[145] C. TAMM, S. PIJUAN GALITÓ, AND C. ANNERÉN, *A comparative study of protocols for mouse embryonic stem cell culturing*, PLOS ONE, 8 (2013), p. e81156.

[146] F. TONA, N. PETSAS, E. SBARDELLA, L. PROSPERINI, M. CARMELLINI, C. POZZILLI, AND P. PANTANO, *Multiple sclerosis: altered thalamic resting-state functional connectivity and its effect on cognitive function*, Radiology, 271 (2014), pp. 814–821.

[147] M. E. TORRES-PADILLA AND I. CHAMBERS, *Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage*, Development, 141 (2014), pp. 2173–2181.

[148] F. TRONARP, N. P. SUBRAMANIYAM, S. SARKKA, AND L. PARKKONEN, *Tracking of dynamic functional connectivity from meg data with kalman filtering*, 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018 (2018), pp. 1003–1006.

[149] N. TZOURIO-MAZOYER, B. LANDEAU, D. PAPATHANASSIOU, F. CRIVELLO, O. ETARD, N. DELCROIX, B. MAZOYER, AND M. JOLIOT, *Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain*, NeuroImage, 15 (2002), pp. 273–289.

[150] N. V. VARLAKHANOVA, R. F. COTTERMAN, W. N. DEVRIES, J. MORGAN, L. R. DONAHUE, S. MURRAY, B. B. KNOWLES, AND P. S. KNOEPFLER, *myc maintains embryonic stem cell pluripotency and self-renewal*, Differentiation, 80 (2010), pp. 9–19.

[151] P. E. VERTES, A. F. ALEXANDER-BLOCH, N. GOGTAY, J. N. GIEDD, J. L. RAPOPORT, AND E. T. BULLMORE, *Simple models of human brain functional networks*, Proceedings of the National Academy of Sciences of the United States of America, 109 (2012), pp. 5868–5873.

[152] S. VOSSEL, J. J. GENG, AND G. R. FINK, *Dorsal and ventral attention systems: distinct neural circuits but collaborative roles*, The Neuroscientist, 20 (2014), pp. 150–159.

[153] J. WANG, D. N. LEVASSEUR, AND S. H. ORKIN, *Requirement of nanog dimerization for stem cell self-renewal and pluripotency*, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008), pp. 6326–6331.

[154] D. J. WATTS AND S. H. STROGATZ, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440–442.

[155] C. WILLYARD, *New human gene tally reignites debate*, Nature, 558 (2018), pp. 354–355.

[156] H. A. WISHART, A. J. SAYKIN, B. C. MCDONALD, A. C. MAMOURIAN, L. A. FLASHMAN, K. R. SCHUSCHU, K. A. RYAN, C. E. FADUL, AND L. H. KASPER, *Brain activation patterns associated with working memory in relapsing-remitting ms*, Neurology, 62 (2004), pp. 234–238.

[157] M. WOJTOWICZ, E. L. MAZEROLLE, V. BHAN, AND J. D. FISK, *Altered functional connectivity and performance variability in relapsing-remitting multiple sclerosis*, Multiple sclerosis Journal, 20 (2014), pp. 1453–1463.

[158] J. WRAY, T. KALKAN, S. GOMEZ-LOPEZ, D. ECKARDT, A. COOK, R. KEMLER, AND A. SMITH, *Inhibition of glycogen synthase kinase-3 alleviates tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation*, Nature Cell Biology, 13 (2011), pp. 838–845.

[159] J. WRAY, T. KALKAN, AND A. G. SMITH, *The ground state of pluripotency*, Biochem Soc Trans, 38 (2010), pp. 1027–1032.

[160] C. W. WU, H. GU, H. LU, E. A. STEIN, J. H. CHEN, AND Y. YANG, *Frequency specificity of functional connectivity in brain networks*, NeuroImage, 42 (2008), pp. 1047–1055.

[161] D. WU AND W. PAN, *Gsk3: a multifaceted kinase in wnt signaling*, Trends in Biochemical Sciences, 35 (2010), pp. 161–168.

[162] R. H. WURTZ AND J. E. ALBANO, *Visual-motor function of the primate superior colliculus*, Annual Review of Neuroscience, 3 (1980), pp. 189–226.

[163] H. XIE, C. Y. ZHENG, D. A. HANDWERKER, P. A. BANDETTINI, V. D. CALHOUN, S. MITRA, AND J. GONZALEZ-CASTILLO, *Efficacy of different dynamic functional connectivity methods to capture cognitively relevant information*, NeuroImage, 188 (2018), pp. 502–514.

[164] M. YAMAJI, J. UEDA, K. HAYASHI, H. OHTA, Y. YABUTA, K. KURIMOTO, R. NAKATO, Y. YAMADA, K. SHIRAHIGE, AND M. SAITOU, *Prdm14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells*, Cell Stem Cell, 12 (2013), pp. 368–382.

[165] F. YI, L. PEREIRA, J. A. HOFFMAN, B. R. SHY, C. M. YUEN, D. R. LIU, AND B. J. MERRILL, *Opposing effects of tcf3 and tcf1 control wnt stimulation of embryonic stem cell self-renewal*, Nature Cell Biology, 13 (2011), pp. 762–770.

[166] P. ZARCHAN AND H. MUSOFF, *Fundamentals of Kalman Filtering: A Practical Approach*, American Institute of Aeronautics and Astronautics, Inc., 2013.

[167] B. ZHANG AND S. HORVATH, *A general framework for weighted gene co-expression network analysis*, Statistical Applications in Genetics and Molecular Biology, 4 (2005), p. Article17.

[168] X. ZHANG, K. A. PETERSON, X. S. LIU, A. P. MCMAHON, AND S. OHBA, *Gene regulatory networks mediating canonical wnt signal-directed control of pluripotency and differentiation in embryo stem cells*, Stem Cells, 31 (2013), pp. 2667–2679.

[169] X. ZHANG, F. ZHANG, D. HUANG, L. WU, L. MA, H. LIU, Y. ZHAO, S. YU, AND J. SHI, *Contribution of gray and white matter abnormalities to cognitive impairment in multiple sclerosis*, International Journal of Molecular Sciences, 18 (2016), p. 46.

[170] F. ZHOU, Y. ZHUANG, H. GONG, B. WANG, X. WANG, Q. CHEN, L. WU, AND H. WAN, *Altered inter-subregion connectivity of the default mode network in relapsing remitting*

multiple sclerosis: a functional and structural connectivity study, PLOS ONE, 9 (2014), p. e101198.

[171] F. ZHOU, Y. ZHUANG, L. WU, N. ZHANG, X. ZENG, H. GONG, AND C. S. ZEE, *Increased thalamic intrinsic oscillation amplitude in relapsing-remitting multiple sclerosis associated with the slowed cognitive processing*, Clinical Imaging, 38 (2014), pp. 605–610.

[172] Y. ZHUANG, F. ZHOU, AND H. GONG, *Intrinsic functional plasticity of the sensorimotor network in relapsing-remitting multiple sclerosis: Evidence from a centrality analysis*, PLOS ONE, 10 (2015), p. e0130524.